



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA MATEMÁTICA

DESARROLLO DE UN MODELO EPIDEMIOLOGICO LAGRANGIANO MULTICLASE
DE TIEMPOS DE RESIDENCIA Y RIESGOS EN AMBIENTES, APLICADO AL
BROTE DE COVID-19 EN LA REGIÓN METROPOLITANA DE CHILE

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS DE LA
INGENIERÍA, MENCIÓN MATEMÁTICAS APLICADAS

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL MATEMÁTICA

TABITA ANDREA CATALÁN MUÑOZ

PROFESOR GUÍA:
AXEL OSSES ALVARADO

MIEMBROS DE LA COMISIÓN:
HÉCTOR RAMÍREZ CABRERA
CARLA CASTILLO LABORDE
PEDRO GAJARDO ADARO

Este trabajo ha sido parcialmente financiado por Proyecto Fondecyt 1191903,
CMM ANID PIA AFB170001, CMM ANID BASAL ACE210010
y CMM ANID BASAL FB210005

SANTIAGO DE CHILE
2022

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGENIERA CIVIL MATEMÁTICA Y
GRADO DE MAGÍSTER EN CIENCIAS DE LA INGENIERÍA,
MENCIÓN MATEMÁTICAS APLICADAS
POR: TABITA ANDREA CATALÁN MUÑOZ
FECHA: 2022
PROF. GUÍA: AXEL OSSES ALVARADO

DESARROLLO DE UN MODELO EPIDEMIOLÓGICO LAGRANGIANO MULTICLASE DE TIEMPOS DE RESIDENCIA Y RIESGOS EN AMBIENTES, APLICADO AL BROTE DE COVID-19 EN LA REGIÓN METROPOLITANA DE CHILE

La pandemia de COVID-19 ha sido un desafío sanitario a nivel mundial desde 2020, causando más de 6 millones de muertes hacia marzo de 2022. Esta enfermedad se contagia por gotitas respiratorias y contacto estrecho, y es por esto que una de las medidas de mitigación más importantes antes del desarrollo de una vacuna fue el distanciamiento social. Sin embargo, medidas como las cuarentenas no han podido ser adoptadas de manera uniforme entre la población y los grupos socioeconómicos más bajos se han visto más afectados.

Se considera una noción de riesgo basada en tres factores; amenaza, exposición y vulnerabilidad. Se busca un modelo compartimental lagrangiano de clases y ambientes que incluya estos tres factores, con el objetivo de estimar la vulnerabilidad, que llamamos factor sanitario, para distintos grupos socioeconómicos.

Se plantea la siguiente metodología: en primer lugar, estudiar los datos disponibles y definir las clases y ambientes a utilizar. A continuación, estimar una matriz de tiempos de residencia que codifique el tiempo que pasa cada clase en cada ambiente, y cómo ha variado a lo largo de la pandemia. En tercer lugar, estimar el factor sanitario para ese modelo a partir de observaciones de datos epidemiológicos reales, mediante la técnica de Filtro de Kalman.

Esta metodología se aplica a un caso de estudio: el desarrollo de la pandemia de COVID-19 en Santiago de Chile. Se estima una matriz de tiempos de residencia utilizando datos de movilidad. Posteriormente se usa Filtro de Kalman Extendido con estado aumentado en un modelo SEIR, y se estima el factor sanitario para cinco grupos socioeconómicos de Santiago, formados por comunas agrupadas según su Índice de Prioridad social. Se utilizan como observaciones el número de casos confirmados por comuna.

Se observan diferencias importantes en los factores sanitarios calculados para los distintos grupos socioeconómicos, lo que apoya que las clases más vulnerables fueran sometidas a condiciones más riesgosas, permaneciendo en ambientes con menor cuidado. Destaca la brusca reducción del factor sanitario estimado para el grupo socioeconómico alto a finales de mayo de 2021. Los casos hipotéticos estudiados apoyan la importancia de las medidas de higiene y cuidado.

La metodología propuesta puede ser fácilmente extendida a otros modelos epidemiológicos compartimentales, pero en el presente trabajo solo se pudo considerar los ambientes hogar y exterior, por lo que no se pudo estudiar el efecto en otros ambientes.

*Pero benditos son los que confían en el Señor
y han hecho que el Señor sea su esperanza y confianza.*
Jeremías 17:7

Agradecimientos

Quiero comenzar agradeciendo a mi familia; no estaría aquí ahora de no ser por ellos. A mi tía Jeanette, porque su corazón limpio, su fe sincera y su amor fiel siguen conmigo aunque ella ya no esté. A mi tío Pedro, porque siempre puedo contar con él. Mi hermano Ángel, porque nadie se alegra más que él cuando llego a la casa. A mi abuelito Pedro, que ya no está, y mi abuelita Lucila, por todo su cariño. A mi mamá Elisa, una fuerza de la naturaleza, porque está dispuesta a confrontarme, porque un abrazo suyo hace que todo se sienta más fácil, y porque no sé qué sería de mí sin ella.

No puedo dejar de mencionar a Sonia Díaz, quien me dio un hogar estos siete años de universidad y es prácticamente familia. Le agradezco por los muchos buenos momentos compartidos, por su preocupación y cuidado, por sus oraciones y también por la paciencia y el llamado de atención cuando lo necesitaba.

Escribir esta tesis ha sido un proceso lleno de altibajos y no habría sido posible sin las personas que me han ayudado en el camino. Agradezco a mi profesor guía Axel Osses, por su apoyo en cada etapa, por su paciencia y porque siempre parece creer en mí más que yo misma. Agradezco también a quienes trabajan en el Laboratorio de Escritura Armadillo, especialmente a Amparo Galdames, que fue un rayo de luz en los días más oscuros, y Enrique Sologuren. Ambos me dieron el mapa para orientarme en el caos de la tesis y cambiaron para siempre mi perspectiva del proceso de investigación.

A mi familia en Cristo, mi iglesia Estanque de Siloé, quienes me ayudan a mirarlo a Él y a esperar en Él con paciencia, e hicieron más llevadero el proceso de escribir una tesis en medio de una pandemia. A Fernando Espinoza e Isaac Jara, por su preocupación y cuidado. A Jamilette Becerra, por las risas. A Víctor Harvey, que el Señor lo tenga en Su gloria, y Victoria Harvey, por las horas de buenas conversaciones. A Maribel Troncoso y Gamaliel Harvey; porque su cariño animó muchos días tristes.

Me siento agradecida por las personas maravillosas que conocí en mi transcurso por la universidad. Evelyn Lorca, una mujer admirable y de las más fuertes que conozco, y una buena amiga. Los de la oficina 435, que me adoptaron y alegraron mis días en el DIM. Destaco al inigualable Felipe Matus, con quien tuve el gusto de trabajar como auxiliar, y a Manuel Suil que fue una fuente de memes y un invaluable apoyo a la hora de lidiar con las frustraciones de la tesis. Los profesores y funcionarios del DIM. Los profesores del Área de Humanidades. La miss Caro Santander, que hizo más felices los cinco semestres en que fui su alumna en inglés.

Finalmente, agradezco a quienes creyeron en mí y me hicieron ver que era posible llegar a la universidad. A Mirta Rivas, que me ha animado todos estos años. A mis profesores, especialmente a Luis Castro, quien me dio el empujoncito que terminaría llevándome a transitar el maravilloso y desafiante camino de aprender matemática, y a Daniela Lizama, una persona increíblemente inspiradora, y quien creyó en mí y se la jugó antes que nadie.

Tabla de Contenido

Introducción	1
1. Marco teórico	7
1.1. Modelos epidemiológicos	7
1.1.1. Modelos compartimentales SIR y SEIR	8
1.1.2. Enfoques euleriano y lagrangiano	9
1.1.3. Modelo multiclase con dispersión virtual	11
1.2. Filtro de Kalman	13
1.2.1. Filtro de Kalman lineal discreto	13
1.2.2. Filtro de Kalman Extendido (EKF)	15
1.2.3. Estimación de parámetros: estado aumentado	16
1.2.4. Suavizado: <i>RTS smoother</i>	17
1.2.5. Restricciones: observaciones (casi) perfectas	18
1.3. Antecedentes: El caso del COVID-19	19
1.3.1. El brote de COVID-19 en Santiago de Chile	19
2. Metodología	21
2.1. Datos disponibles	22
2.1.1. Censo 2017	22
2.1.2. Índice de prioridad social (IPS) 2019	23
2.1.3. Encuesta Origen-Destino Santiago 2012	23
2.1.4. Informes de Movilidad Local sobre COVID-19	24

2.1.5.	Uso de infraestructura de telecomunicaciones	24
2.1.6.	Datos-COVID19	25
2.2.	Decisiones específicas acerca del modelo	27
2.3.	Matriz de tiempos de residencia	28
2.3.1.	Elección de clases y ambientes	29
2.3.2.	Matriz de Santiago en condiciones normales	30
2.3.3.	Variaciones de la matriz a lo largo del tiempo	32
2.4.	Estimación de parámetros	33
2.4.1.	Elección de Filtro de Kalman	33
2.4.2.	Estimación de la tasa de contagios	35
2.4.3.	Estimación de parámetros fijos	39
2.4.4.	Matrices de covarianza: ajuste por caso sintético	39
2.4.5.	Observabilidad	40
2.4.6.	Implementación	40
2.5.	Evaluación del modelo mediante casos hipotéticos	41
3.	Resultados	42
3.1.	Matriz de tiempos de residencia	42
3.1.1.	Santiago en condiciones normales	43
3.1.2.	Santiago a lo largo del tiempo	46
3.2.	Estimación de parámetros	47
3.2.1.	Modelo multiclase con datos sintéticos	47
3.2.2.	Sensibilidad con respecto a β_{exterior} y α_0	50
3.2.3.	Sensibilidad ante la covarianza	51
3.2.4.	Modelo con datos reales	52
3.2.5.	Sensibilidad ante los parámetros γ_E, γ_I	56
3.3.	Evaluación del modelo mediante casos hipotéticos	57
3.3.1.	Escenario con contagio exterior predominante	57

3.3.2. Escenario con contagio en hogar	59
4. Discusión	62
4.1. Acerca del caso de estudio	62
4.2. Acerca de la metodología	64
Conclusión	66
Bibliografía	69
Anexo: Clasificación IPS 2019	76

Índice de Tablas

2.1. Resumen de las categorías socioeconómicas por IPS, dadas por [1].	30
3.1. Criterios usados para obtener las clases de la matriz detallada a partir de la EOD2012 Santiago.	43
3.2. Propósitos de viaje de la EOD2012 y sus ambientes asociados.	44
3.3. Modos de transporte de la EOD2012 y sus ambientes asociados.	44
3.4. Matriz de tiempos de residencia para las simulaciones.	46
3.5. Algunas fechas relevantes para el desarrollo de la pandemia en Santiago. . .	54
3.6. Casos hipotéticos para distintas combinaciones de cuarentena y cuidado. . .	57
3.7. Observaciones para cada caso hipotético, escenario con $\beta_{\text{exterior}} = 68$	58
3.8. Observaciones para cada caso hipotético, escenario con $\beta_{\text{exterior}} = 1.2$	60
1. Clasificación de las comunas de la RM por IPS 2019	76

Índice de Ilustraciones

1.	Diferencias en la variación de movilidad de cinco comunas de la Región Metropolitana.	4
1.1.	Diagrama de flujo del modelo SIR dado por las ecuaciones 1.1.	8
1.2.	Diagrama de flujo del modelo SEIR dado por las ecuaciones 1.2.	9
2.1.	Variación en la movilidad agregada por actividad en la Región Metropolitana y media móvil de 7 días. Fuente: Elaboración propia con datos de Google. . .	24
2.2.	Variación en la movilidad de seis comunas de la Región Metropolitana. Fuente: ISCI Covid Analytics.	25
2.3.	Total de casos fallecidos por COVID-19 acumulados a nivel nacional. Fuente: Centro de Modelamiento Matemático con Datos-COVID19.	26
2.4.	Porcentaje de cobertura vacunación COVID-19 en la Región Metropolitana. Fuente: Centro de Modelamiento Matemático con Datos-COVID19.	26
2.5.	Transformar propósitos de viaje en tiempo de actividad.	32
3.1.	Matriz detallada de tiempos de residencia para Santiago	45
3.2.	Fracción de la movilidad en pandemia por clase, c/r a movilidad normal. . .	46
3.3.	Tiempo de residencia en exterior para las 5 clases elegidas.	47
3.4.	Ejemplo de estimación de estados y parámetros variables en el tiempo con <i>Kalman Smoother</i> , a partir de las observaciones del número de casos acumulados C_1, C_2	48
3.5.	Número de Infectados, comparando los resultados obtenidos con la solución real, en valores absolutos y normalizados con respecto a la cantidad de personas por clase.	49
3.6.	Factor sanitario, caso sintético	49
3.7.	Sensibilidad ante riesgo exterior y condición inicial del factor sanitario. . . .	50

3.8. Sensibilidad ante la varianza y condición inicial del factor sanitario.	51
3.9. <i>Kalman RTS Smoother</i> aplicado al caso real	52
3.10. Cantidad de Infectados estimados a partir de datos reales, en valores absolutos y normalizados con respecto a la cantidad de personas por clase.	53
3.11. Cantidad de Susceptibles estimados a partir de datos reales, en valores absolutos y normalizados con respecto a la cantidad de personas por clase.	54
3.12. Factor sanitario estimado a partir de datos reales.	55
3.13. Sensibilidad factor sanitario estimado para los datos reales, ante distintos γ_I	56
3.14. Personas infectadas para casos hipotéticos seleccionados. Escenario $\beta_{\text{exterior}} \gg 1$	59
3.15. Personas infectadas para caso hipotéticos seleccionados.	61

Introducción

Motivación y antecedentes

El COVID-19 es una enfermedad provocada por el virus SARS-CoV-2. Fue descubierta a finales de 2019, se convirtió en pandemia en 2020 y ha sido un desafío sanitario a nivel mundial; hacia marzo de 2022 ya ha causado más de 6 millones de muertes en todo el mundo según datos de la Universidad John Hopkins [2].

El COVID-19 se transmite por microgotas emitidas al respirar, hablar, toser o estornudar, o por contacto estrecho [2][3]. Se ha intentado mitigar mediante diferentes medidas [4][5], que incluyen varias formas de distanciamiento social como reducciones de movilidad voluntarias, cuarentenas totales o parciales, teletrabajo, distancias mínimas entre personas, etc. Se han fomentado además distintas medidas de higiene como el uso de mascarillas, lavado frecuente de manos, ventilación de espacios, entre otras.

Con respecto a las medidas farmacéuticas, durante 2020 se estudiaron y desarrollaron varios intentos de vacunas, utilizando conocimientos adquiridos en la lucha contra virus como SARS-CoV y MERS-CoV. En 2021 se comenzó la vacunación masiva y hacia marzo de 2022, más de la mitad de la población mundial ha recibido al menos una dosis, con países como China y Singapur teniendo más de 85% de su población vacunada según datos publicados en *Our World in Data* [6].

El impacto de la enfermedad ha sido heterogéneo entre la población. Se sabe que la edad es un factor a considerar; la susceptibilidad a la infección y la probabilidad de desarrollar síntomas luego del contagio aumentan con la edad [7]. Similarmente, tanto el porcentaje de infectados cuya gravedad requiere hospitalización como el porcentaje que termina falleciendo es más alto entre los adultos mayores [8]. El nivel socioeconómico tampoco puede ser ignorado; el menor acceso a la salud [9], sumado al desempleo y la existencia de enfermedades crónicas previas [10] son un agravante que hace más vulnerables a los niveles socioeconómicos más pobres.

Los datos de movilidad han sido utilizados ampliamente para modelar el avance de la enfermedad [11][12][13]; en la mayoría de los países la movilidad explica una parte importante de la variaciones en la transmisibilidad [14]. La referencia [15] nota que las diferencias en movilidad explican las diferencias en transmisión en diferentes grupos económicos, notando que además las clases de menor nivel socioeconómico se enfrentan a ambientes más riesgosos.

El Grupo Intergubernamental de Expertos sobre el Cambio Climático (*Intergovernmental Panel on Climate Change*, también conocido como IPCC) define [16] el riesgo como una combinación entre tres factores: amenaza, exposición y vulnerabilidad. La amenaza se refiere a la posible ocurrencia de un evento que podría tener efectos adversos en los elementos expuestos y vulnerables. La exposición se refiere al hecho de estar presente en el lugar donde ocurre ese evento. La vulnerabilidad se refiere a qué tan propensos son los elementos expuestos a sufrir efectos adversos al ser impactados por una amenaza, y se relaciona con su predisposición y fragilidad.

Lo anterior motiva la siguiente idea: estudiar el riesgo de contagio de COVID-19 al que se enfrentan los distintos grupos socioeconómicos en diversos ambientes, utilizando un modelo que considere estos tres factores. ¿En qué lugares pasan su tiempo? ¿Cuántas amenazas hay en esos lugares? ¿Cuánto tiempo pasan expuestos a esas amenazas? ¿Qué tan vulnerables son a esas amenazas?

Algunos lugares interesantes a considerar son el hogar, el trabajo, la escuela, el transporte público, entre otros. La amenaza sería la posibilidad de ser contagiado en cierto lugar, lo que está relacionado a la cantidad de personas presentes en dicho lugar, cuántas de ellas están infectadas y a las características del lugar en sí; no es lo mismo estar en el transporte público que en un parque. La vulnerabilidad es alguna característica del grupo que le hace estar peor preparado ante la amenaza, por ejemplo, la falta o mala aplicación de medidas de cuidado como uso de la mascarilla, lavado de manos, mantenerse a una distancia mínima de los demás, etc. También incluye la susceptibilidad del grupo al contagio.

Un modelo epidemiológico compartimental separa a la población en grupos o compartimientos de acuerdo al grado de avance de la enfermedad, y asigna reglas para pasar de un compartimiento a otro. Se busca un modelo epidemiológico compartimental que considere los factores antes mencionados. Suponiendo que hay varias clases $i \in 1 \dots n$ y ambientes $j \in 1 \dots m$, se propone una idea de amenaza de la forma $\beta_j I_j / N_j$, que incluye la fracción de infectados I_j / N_j en un lugar j , y un factor β_j dependiente de las características del lugar. Se consideran además valores p_{ij} , que representan la fracción de tiempo que la clase i está en el ambiente j , y un factor de vulnerabilidad α_i dependiente de la clase i . Al combinar lo anterior se obtiene la expresión 1, que da una idea de la tasa de contagios para la clase i buscada.

$$\alpha_i \sum_{j=1}^m \beta_j p_{ij} \frac{I_j}{N_j} \quad (1)$$

El modelo propuesto es una variante del modelo de dispersión virtual presentado por [17] y desarrollado posteriormente en [18] como un *framework* para incorporar clases y ambientes en un modelo epidemiológico. Este enfoque tiene dos ventajas que lo hacen interesante: por una parte permite trabajar con heterogeneidad espacial y en clases simultáneamente. Pero además de eso utiliza los tiempos de residencia p_{ij} y el riesgo β_j del ambiente como un *proxy* del número de contactos efectivos.

Si bien la noción de contactos efectivos es clara para enfermedades de transmisión sexual o de enfermedades transmitidas por vectores, para el caso de enfermedades transmitidas

por contacto estrecho la noción es mucho más vaga y difícil de definir. Suele modelarse con matrices del tipo WAIFW (*Who acquires infection from whom?* o ¿Quién adquiere la infección de quién?), que a su vez son aproximadas por matrices de mezcla social o *social mixing* [19][20]. Este método ha sido aplicado al COVID-19 [21].

[17] y [18] utilizan un enfoque diferente; en lugar de intentar estimar quién tiene contacto con quién, las distintas clases interactúan en ambientes, de forma que no se cuenta cuántas veces una clase interactúa con otra, sino que se busca saber cuánto tiempo pasa cada clase en un ambiente. A este tiempo se le llamará tiempo de residencia. Los ambientes visitados por cada grupo y el tiempo pasado ahí se asocia directamente con la movilidad, de hecho, la relación entre viajes y uso del tiempo es ampliamente utilizada; existen modelos [22][23] de predicción de demanda de viajes que utilizan un enfoque basado en actividades, y [24] obtiene tiempos de actividad a partir del registro de viajes de Encuestas Origen-Destino.

Puesto que la movilidad y los tiempos de residencia pueden considerarse conocidos, y que los riesgos β_j pueden ser aproximados, se busca estimar la vulnerabilidad al contagio α_i , que en este modelo está desacoplada de la movilidad, la llamaremos de ahora en adelante “factor sanitario”, puesto que se relaciona con las distintas medidas de distanciamiento social no relacionadas a la movilidad, como lo son el uso de mascarillas, lavado de manos, cumplimiento de la distancia mínima entre personas, entre otras.

Consideraremos que el factor sanitario es variable en el tiempo. Para estimar parámetros, una técnica común es Mínimos Cuadrados o Mínimos Cuadrados Recursivo (RLS) [25][26]. Una alternativa que permite incorporar más información de la dinámica del sistema es utilizar Filtro de Kalman [27], una técnica ampliamente utilizada en ingeniería [28], y que también se ha usado en algunos modelos de COVID-19 [29][30][25].

La metodología será puesta a prueba estudiando el desarrollo de la pandemia por COVID-19 en la ciudad de Santiago de Chile. Ya desde el comienzo de la pandemia en Santiago, [31] hacía notar las dificultades de las clases socioeconómicas más bajas para cumplir las cuarentenas. Esto se observa claramente en la figura 1, donde a lo largo de todo el período mostrado la comuna de Las Condes, una de las más acomodadas, tiene una reducción de movilidad superior a las demás comunas, mientras que La Pintana, una de las comunas con menos recursos, se mantiene con una reducción de movilidad inferior. Varios estudios [32][33][34] han notado las significativas diferencias en el impacto de la pandemia en los distintos sectores socioeconómicos de la capital del país, atribuyéndolas, entre otros factores, a la capacidad de cumplimiento de las cuarentenas y las diferencias en el acceso a salud.

Además de lo anterior, el Ministerio de Ciencia, Tecnología, Conocimiento e Innovación de Chile ha hecho disponibles públicamente [35] varias series de datos epidemiológicos de casos infectados confirmados, hospitalizados, hospitalizados UCI, fallecidos, vacunados, etc. Estos datos poseen diversos niveles de segregación, por comuna, edad o sexo. Todas estas características hacen de la ciudad de Santiago un buen escenario donde implementar el modelo.

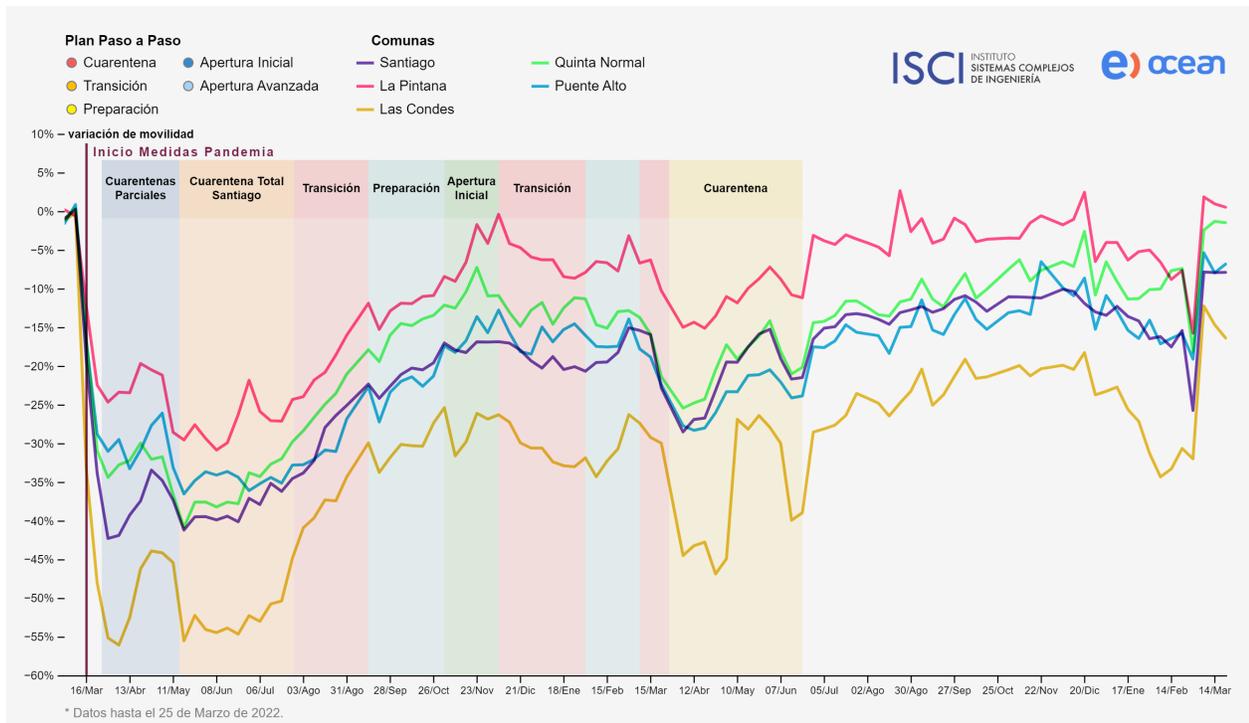


Figura 1: Diferencias en la variación de movilidad de cinco comunas de la Región Metropolitana, desde marzo de 2020 a marzo de 2022. Fuente: ISCI Covid Analytics.

Objetivos

El objetivo principal del trabajo es estimar el factor sanitario de distintas clases sociales, por medio del *framework* lagrangiano de clases y ambientes de [18]. Este objetivo se descompone en dos objetivos específicos. En primer lugar, se necesita definir una metodología que permita estimar el factor sanitario. En segundo lugar, se debe implementar y evaluar la metodología, aplicándola al caso de estudio: la pandemia de COVID-19 en Santiago.

Definir la metodología requiere plantear el modelo lagrangiano, haciéndole los ajustes necesarios al planteado por [18] y considerando las características del COVID-19. Se debe además definir cómo estimar la matriz de tiempos de residencia en base a los datos disponibles. Finalmente, es necesario elegir la variante de filtro de Kalman específica a utilizar en la estimación de parámetros. La implementación requiere de la obtención de datos, el cálculo de la matriz de tiempos de residencia, y la escritura del código para el modelo y el filtro de Kalman.

Metodología

Para cumplir los objetivos antes expuestos se lleva a cabo la siguiente metodología, la cual se divide en varias etapas. En primer lugar, es necesario explorar las distintas fuentes de datos disponibles, tanto de movilidad y viajes como las series de tiempo epidemiológicas, como cantidad de casos registrados, fallecidos, etc. Esto permite decidir el modelo específico

a utilizar, lo que incluye la elección de compartimientos a utilizar (susceptibles, expuestos, infectados, etc), parámetros, además de las clases y ambientes. Todo esto debe considerar las características del COVID-19 y también los datos disponibles.

Una vez decididas las clases y ambientes, es necesario estimar una matriz de tiempos de residencia, que diga cuánto tiempo pasa cada clase en cada lugar. Se desea que esta matriz sea variable en el tiempo, de forma que incorpore cambios en la movilidad causados por las medidas de mitigación, como cuarentenas, teletrabajo, etc.

Todo lo anterior ya permite correr un modelo de la ciudad de Santiago, definiendo factores sanitarios sintéticos y obtener simulaciones. Se busca ahora estimar el factor sanitario a partir de observaciones del modelo. Para esto se utiliza Filtro de Kalman [27], y es necesario elegir e implementar una de sus muchas variantes [36][37] (extendido, *unscented*, filtro por ensambles, etc), y elegir cómo usar el filtro para hacer estimación de parámetros (estado aumentado, filtro de Kalman con múltiples modelos, etc).

Posteriormente se estiman los factores sanitarios para cada una de las clases elegidas, ajustando el modelo a los datos reales de la ciudad de Santiago. Para esto es necesario definir qué variables del modelo observar, que datos utilizar como observación y cómo ajustar los parámetros específicos del filtro.

Finalmente, se utiliza la matriz de tiempos de residencia y los factores sanitarios obtenidos para generar varios casos hipotéticos; distintas variaciones en cuarentenas, diferentes niveles de cuidado. Esto entrega varios desarrollos hipotéticos de la pandemia, y se usará para evaluar el modelo.

Contribuciones y trabajos relacionados

A continuación se describen las principales contribuciones de este trabajo. En primer lugar, la propuesta de una metodología que permite estudiar las medidas de cuidado de forma independiente a la movilidad. En segundo lugar, la aplicación del *framework* lagrangiano de clases y ambientes presentado por [18] a un caso de estudio, lo cual no había sido hecho con anterioridad. Este trabajo guarda, sin embargo, cierta similitud con [38], que plantea un modelo que incorpora patrones de actividad.

En tercer lugar, la implementación en Matlab de la técnica propuesta por [24] de obtención de uso del tiempo partir de los viajes de una Encuesta Origen-Destino. Esta implementación es de código abierto y se encuentra isponible en el repositorio <https://github.com/tabitaCatalan/lagrangian-time>. Se utilizan datos de la Encuesta Origen-Destino Santiago 2012, puesto que la encuesta se realiza cada 10 años y su versión 2022 no está aún disponible.

En cuarto lugar, el desarrollo de una librería, escrita en el lenguaje Julia, para el trabajo con Filtro de Kalman con dinámica y observadores lineales y/o no lineales. Esta se encuentra documentada, es extensible y de código abierto. Está disponible en el repositorio <https://github.com/tabitaCatalan/kalman>. Existen librerías similares en el mismo lenguaje como

KalmanFilters.jl y Kalman.jl, pero estas no fueron usadas debido a que en la etapa de decidir qué Filtro utilizar, ninguna contaba con todas las opciones que se quería explorar.

En quinto lugar, el planteamiento de una metodología para el ajuste de los hiperparámetros relacionados a la covarianza del filtro de Kalman, al estimar parámetros de un modelo epidemiológico. Este era un vacío en trabajos como [29] y [25], donde se limitan a elegir un valor fijo para el caso particular que trabajan, sin ofrecer una justificación.

Finalmente, la aplicación del modelo al desarrollo de la pandemia de COVID-19 en la ciudad de Santiago de Chile, lo que permite una estimación de la vulnerabilidad/factor sanitario, de forma independiente a la movilidad. Modelos anteriores de Santiago como [34] solo consideran movilidad.

Estructura de la tesis

El resto de esta tesis está organizado como sigue. El capítulo 1 establece las bases para entender el trabajo realizado. Se presentan varios conceptos relevantes de modelos epidemiológicos y de la teoría de filtro de Kalman. Se exponen también los antecedentes del caso de estudio; la enfermedad COVID-19 y su desarrollo en la ciudad de Santiago de Chile.

El capítulo 2 justifica y detalla los procedimientos seguidos; cómo se usó el framework de [18], los datos utilizados, los supuestos y simplificaciones hechas, el modelo final utilizado y las técnicas específicas implementadas.

El capítulo 3 presenta los principales resultados obtenidos; la matriz de tiempos de residencia, el caso sintético utilizado para ajustar los parámetros del filtro de Kalman y las estimaciones para el factor sanitario obtenidas utilizando datos reales. Se muestran también los diferentes escenarios hipotéticos generados a partir de esas estimaciones.

El capítulo 4 discute los resultados obtenidos para el caso de estudio, sus implicaciones y limitaciones. Comenta además el *framework* lagrangiano planteado en [18], sus ventajas y dificultades a la hora de implementarlo. Se plantean posibles extensiones y trabajo futuro.

Capítulo 1

Marco teórico

En este apartado se presentarán las bases teóricas y conceptuales que sustentan el trabajo realizado. Se presentan también los antecedentes del caso de estudio a tratar: el desarrollo de la pandemia en la Región Metropolitana de Santiago de Chile.

Este capítulo se organiza como sigue; en primer lugar, en la sección 1.1, se presentan los modelos epidemiológicos compartimentales deterministas. Se explican dos enfoques que permiten tratar con poblaciones heterogéneas: lagrangiano y euleriano, y se presenta el modelo lagrangiano propuesto en [18] que se usará en este trabajo.

La sección 1.2 está dedicada a la teoría del Filtro de Kalman, una técnica que permite la estimación de estados en sistemas lineales y que ha sido ampliamente extendida y utilizada, contando con aplicaciones en diversos ámbitos de ingeniería y ciencias como climatología, control de vehículos autónomos, etc. Se presentará la teoría básica del filtro y las principales extensiones que se utilizan durante el trabajo.

Contando ya con las bases teóricas, la sección 1.3 presenta los antecedentes relevantes al caso de estudio a tratar; el brote de COVID-19 en la ciudad de Santiago de Chile. Se describen características particulares de esta enfermedad y cómo influyen en el modelamiento. Se presentan características de la ciudad de Santiago que han influido el desarrollo de la pandemia y las medidas tomadas para ralentizar su avance.

1.1. Modelos epidemiológicos

Se presentan ahora los modelos epidemiológicos compartimentales deterministas. Primero, 1.1.1 explica brevemente los modelos SIR y SEIR. A continuación, 1.1.2 explica los enfoques lagrangiano y euleriano, que permiten tratar con poblaciones heterogéneas. Finalmente, 1.1.3 presenta el modelo lagrangiano propuesto en [18] que se usará en este trabajo.

1.1.1. Modelos compartimentales SIR y SEIR

Los modelos compartimentales son ampliamente utilizados a la hora de modelar una enfermedad contagiosa. La idea tras ellos es simple: separar a la población en grupos o compartimientos de acuerdo al grado de avance de la enfermedad, y asignar reglas para pasar de un compartimiento a otro.

Un modelo muy sencillo es el **SIR**, que separa a la población en **S**usceptibles, **I**nfectados y **R**emovidos. En este trabajo se trata este modelo de manera determinista, viéndolo como el sistema de ecuaciones diferenciales ordinarias 1.1, pero también es posible tratarlo de manera estocástica [39].

$$\begin{aligned} S'(t) &= -\alpha \frac{S(t)I(t)}{N} \\ I'(t) &= \alpha \frac{S(t)I(t)}{N} - \gamma_I I(t) \\ R'(t) &= \gamma_I I(t). \end{aligned} \tag{1.1}$$

Los susceptibles son el grupo de la población que puede ser contagiado con la enfermedad. La enfermedad solo es propagada por los individuos infectados. Cuando un susceptible se infecta, inmediatamente comienza a contagiar a otros. Tras un tiempo los infectados dejan de contagiar y son removidos. Esto puede deberse a una recuperación de la enfermedad con inmunidad completa (no pueden volver a contagiarse), al aislamiento del contagiado o bien a su muerte debido a la enfermedad. El diagrama de flujo para pasar de un compartimiento a otro puede verse en la figura 1.1.

En este modelo, el total de la población dado por $N = S + I + R$ permanece constante a lo largo del tiempo. Este modelo supone que un individuo hace en promedio α contactos “efectivos” o “suficientes” para contagiar la enfermedad por unidad de tiempo. Esto, en conjunto con la probabilidad S/N de que el individuo contactado sea susceptible, deja una cantidad de $\alpha SI/N$ nuevos contagios por unidad de tiempo. Suponer una tasa de recuperación de la forma $\gamma_I I$, con γ_I un valor positivo, implica un periodo de infección con distribución exponencial y media $1/\gamma_I$ (más detalles en [40]).

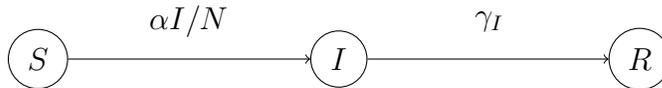


Figura 1.1: Diagrama de flujo del modelo SIR dado por las ecuaciones 1.1.

Es posible complejizar este modelo añadiendo compartimientos extras. El compartimiento de **E**xpuestos permite pasar por un periodo de incubación o latencia, de $1/\gamma_E$ en promedio, antes de comenzar a contagiar la enfermedad, como se ve en las ecuaciones 1.2. Esto da lugar al modelo **SEIR**, cuyo diagrama puede verse en la figura 1.2. Es posible agregar otro compartimientos como hospitalizados, fallecidos, etc. Desde luego, los compartimientos del modelo dependen de la enfermedad a modelar.

$$\begin{aligned}
S' &= -\alpha \frac{SI}{N} \\
E' &= \alpha \frac{SI}{N} - \gamma_E E \\
I' &= \gamma_E E - \gamma_I I \\
R' &= \gamma_I I.
\end{aligned}
\tag{1.2}$$

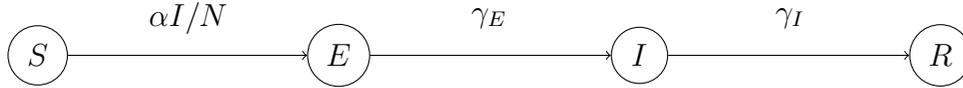


Figura 1.2: Diagrama de flujo del modelo SEIR dado por las ecuaciones 1.2.

1.1.2. Enfoques euleriano y lagrangiano

Los modelos dados por las ecuaciones 1.1 y 1.2 suponen que la población se mezcla perfectamente, de forma que las interacciones entre individuos son homogéneas. Una forma de incluir heterogeneidad es separar a la población en grupos, en un modelo multiclase. Esto permite considerar las distintas capacidades de enfrentar y sobrevivir a la enfermedad de cada grupo, así como las diferencias en comportamiento.

La edad es un factor a considerar en varias enfermedades; niños menores de 5 años son especialmente vulnerables a enfermedades como la malaria o la tuberculosis. Adultos mayores tienen mayor probabilidad de experimentar síntomas graves al contraer influenza. En las enfermedades de transmisión sexual, las parejas tienen edades similares en muchos casos. Para más detalles de cómo considerar la edad dentro del modelamiento epidemiológico puede verse el capítulo 13 de [40].

La ubicación espacial es otro factor a tener en cuenta; en una sociedad interconectada las enfermedades se propagan rápidamente entre un lugar y otro, y cambios en los patrones de movimiento alteran la transmisión de una enfermedad. Este tema es tratado brevemente en el capítulo 14 de [40].

El objetivo de esta subsección es presentar dos enfoques distintivos que se observan al utilizar modelos con mezcla espacial heterogénea; euleriano y lagrangiano. Los nombres derivan de cierta similitud entre estos enfoques y las especificaciones euleriana y lagrangiana del movimiento de fluidos; en el enfoque lagrangiano, es posible “seguir la trayectoria” de las personas/partículas, mientras que en el euleriano, el enfoque está en las regiones y en los flujos de personas/partículas que entran o salen de ahí. Estas definiciones son bastante generales y no son excluyentes entre sí [41].

Se supone una situación con n regiones, donde la personas puede moverse de una región a otra. Por simplicidad se supone además que infección no altera los patrones de movimiento de la población.

Enfoque euleriano

El enfoque euleriano podría describirse como “migratorio”; las personas no pertenecen a ninguna región, sino que se mueven entre ellas. El movimiento de la población entre dos regiones cualesquiera es conocido.

Se presenta ahora un ejemplo más específico, tomado de una simplificación del modelo euleriano de [42]. Se supone conocida una matriz $C = (c_{ij})_{i,j=1\dots n}$ de tasas de migración entre regiones. $c_{ij} \geq 0$ corresponde al número de personas por unidad de tiempo que sale desde la región i hacia la región j . Por definición $c_{ii} = 0$.

Similar a [41], si $N_i(t)$ corresponde a la población presente en la región i a tiempo t , entonces se puede formar un modelo migratorio considerando las personas que salen y las que entran, con las ecuaciones 1.3.

$$N'_i = \sum_{j=1}^n c_{ji} N_j - \sum_{j=1}^n c_{ij} N_i. \quad (1.3)$$

Luego, es posible considerar la propagación de la enfermedad dentro de cada región. Si $S_i(t), I_i(t)$ corresponden al número de susceptibles e infectados en la región i a tiempo t , y α_i es el número promedio de contactos efectivos que hace un individuo en la región i . La dinámica de los susceptibles en este ejemplo puede verse en la ecuación 1.4.

$$S'_i = -\alpha_i \frac{S_i I_i}{N_i}. \quad (1.4)$$

Enfoque lagrangiano

Otra forma de modelar esta situación es enfocarse en las personas; cada una habita en cierta región, aunque puede visitar otras regiones. Esto permite calcular el número de infectados que residen (no están solo de visita) en cada región. Los habitantes de una región usualmente comparten parámetros epidemiológicos (tasa de exposición, recuperación, etc).

Los nuevos contagios se calculan a partir del número de infectados y susceptibles presentes en la región, comparado al total de personas en cada región, que pueden ser visitantes de otras regiones, o habitantes de esa región.

Un ejemplo de modelo lagrangiano se encuentra en [43]. Este define $S_{ij}(t), I_{ij}(t), R_{ij}(t)$ como los susceptibles, infectados y recuperados de la clase i que se encuentran en la clase j a tiempo t . Esto permite calcular, definiendo $N_{ij} = S_{ij} + I_{ij} + R_{ij}$, el total de residentes de la región i como $N_i^r := \sum_{j=1}^n N_{ij}$ y el total de personas presentes en la región i como $N_i^p := \sum_{j=1}^n N_{ji}$. La tasa de infección de la población de la región i debido a los contagios producidos en la región k se calcula simplemente como la suma de todos los contactos efectivos entre susceptibles de la clase i con algún infectado de cualquier clase, como se ve en la ecuación 1.5. α_{ikj} corresponde a la cantidad de contactos efectivos entre la clase i y la clase

j que ocurren en la región k .

$$S'_{ik} = - \sum_{j=1}^n \alpha_{ikj} \frac{S_{ik} I_{jk}}{N_j^p}. \quad (1.5)$$

Este modelo, guardando similitud con el modelo euleriano expuesto antes, supone que los residentes susceptibles de la región i viajan a otras regiones a una tasa constante σ_i . Desde la región i van a la región j con probabilidad ν_{ij} , y vuelven a su región con tasa ρ_{ij} . Esta, sin embargo, no es la única forma de trabajar con la movilidad de la población. Los artículos [41] y [17] utilizan matrices $P = (p_{ij})_{i,j=1\dots n}$ de tiempos de residencia, donde p_{ij} corresponde a la fracción de su tiempo que los residentes de la clase i pasan en la región j . Desde luego se cumple $\sum_{j=1}^n p_{ij} = 1$ para cada i . Esta forma de abordar la movilidad será la más interesante para este trabajo, y se verá en más detalle en la siguiente sección.

1.1.3. Modelo multiclase con dispersión virtual

Si bien separar a la población en distintos grupos permite incorporar heterogeneidad en las interacciones, estos modelos añaden dificultades extra, puesto que ahora es necesario conocer de qué manera interactúan los distintos grupos. La primera dificultad reside en la noción de contacto efectivo. Si bien esta es clara en el contexto de enfermedades de transmisión sexual o enfermedades transmitidas por vectores (como la malaria y el dengue, que son transmitidas por mosquitos), el concepto es mucho más vago al referirnos a enfermedades transmitidas por contacto estrecho o por microgotas al toser, estornudar o hablar.

Otra dificultad está en la necesidad de conocer en gran detalles las interacciones entre grupos. En la sección anterior, el modelo euleriano presentado requería una matriz $C = (c_{ij})_{i,j=1\dots n}$ con las tasas de migración entre cualquier par de regiones. El modelo lagrangiano, similarmente, requería las probabilidades ν_{ij} de pasar de una región a otra, para cualquier par de regiones. Más aún, necesitaba de una matriz α_{ikj} que detallara quién estaba en contacto con quién y dónde.

Este tipo de matrices de contacto es conocida como WAIFW (*Who Acquires Infection From Whom* o ¿Quién adquiere la infección de quién?) [44]. Este tipo de matrices suelen aproximarse por matrices de “mezcla social” o *social mixing*; varios ejemplos para enfermedades transmitidas por contactos estrechos son [19][45][46].

[17] sugiere una solución alternativa; utilizar un modelo lagrangiano con tiempos de residencia. De esta forma, el tiempo pasado en cada región y su riesgo sirve como *proxy* de la probabilidad de contagios. La referencia [17] propone dos variantes de esta idea; en la primera, cada clase pertenece a una región pero pasa un tiempo en las otras regiones. Esta forma de abordar el problema es utilizada también por [41]. En la segunda variante, las clases interactúan en ambientes como hospitales, centros comerciales, escuelas, transporte público, lugares de trabajo, etc. Esto permite combinar clases y dispersión espacial en un mismo modelo. Esta variante es extendida posteriormente en [18] y es la que se usa en este trabajo.

Se separa a la población en n clases, las cuales interactúan en m ambientes o áreas de

riesgo. $N_i(t)$, $i = 1, \dots, n$ corresponde a la población de la clase i . Suponemos que la población de la clase i pasa una proporción $p_{ij} \in [0, 1]$ de su tiempo en el ambiente j , con $\sum_{j=1}^m p_{ij} = 1$ para cada i . Para casos extremos, por ejemplo, podría darse que $p_{ij} = 0$, es decir, la clase i no gasta nada de su tiempo en el ambiente j , mientras que $p_{ij} = 1$ significa que la clase i pasa todo su tiempo en el ambiente j .

Para ejemplificar se usan compartimientos SIR, sin embargo, esta metodología puede extenderse a otros compartimientos. Las ecuaciones en este caso están dadas por 1.6, donde λ_i está dado por 1.7. Los términos b_i y d_i representan las tasas de natalidad y mortalidad de la clase i respectivamente.

$$\begin{aligned} S_i(t)' &= b_i - d_i S_i - \lambda_i(\vec{x}, t) \\ I_i(t)' &= \lambda_i(\vec{x}, t) - \gamma_i I_i(t) - d_i I_i \\ R_i(t)' &= \gamma_i I_i(t). \end{aligned} \tag{1.6}$$

$$\lambda_i(\vec{x}, P) = \sum_{j=1}^m \beta_j p_{ij} S_i(t) \frac{\sum_{k=1}^n p_{kj} I_k}{\sum_{k=1}^n p_{kj} N_k}. \tag{1.7}$$

El valor β_j es una medida de riesgo propio del ambiente j -ésimo y depende de las condiciones ambientales y sanitarias. Representa la idea de que no es lo mismo pasar el tiempo en un parque que en el transporte público o en un bar. Esto debido a factores como la distancia que mantienen las personas en esos ambientes, la ventilación y sanitización a la que son sometidos, entre otros.

Puesto que p_{ij} es la proporción del tiempo que la clase i reside en el ambiente j , entonces la cantidad de personas en promedio de la clase i en el ambiente j , en algún momento de tiempo t , está dada por $N_i(t)p_{ij} = S_i(t)p_{ij} + I_i(t)p_{ij} + R_i(t)p_{ij}$. La población total en el ambiente j es $\sum_{k=1}^m N_k p_{ij}$, de la cual $\sum_{k=1}^m I_k p_{ij}$ están infectados. De esta forma, el número de infectados de la clase i en el ambiente j por unidad de tiempo está dado por la ecuación 1.8.

$$\underbrace{b_j}_{\text{riesgo}} \cdot \underbrace{S_i p_{ij}}_{\substack{\text{susceptibles} \\ \text{de la clase } i \\ \text{en el ambiente } j}} \cdot \underbrace{\frac{\sum_{k=1}^m I_k p_{ij}}{\sum_{k=1}^m N_k p_{ij}}}_{\substack{\text{proporción} \\ \text{de infectados} \\ \text{en el ambiente } j}} \tag{1.8}$$

El modelo presentado supone que la etapa de la enfermedad no afecta el comportamiento de la población. Una forma más completa de modelarlo es seguir a [18] y utilizar distintas matrices de tiempos de residencia para cada compartimiento, representando ideas como que los infectados se queden en su casa guardando cuarentena por ejemplo.

1.2. Filtro de Kalman

Filtro de Kalman es un método ampliamente utilizado en diversas aplicaciones de ingeniería como navegación, control de vehículos, procesamiento de señales, etc [28]. Se tiene un sistema cuya evolución está dada por ciertas ecuaciones. El estado del sistema es desconocido pero se cuenta con observaciones ruidosas de este. El objetivo es estimar el estado a partir de estas observaciones; por ejemplo, estimar la posición de un vehículo cuando solo se conoce su velocidad.

Para un instante k -ésimo, hay tres formas importantes de estimación del estado x_k del sistema. El suavizado (*smoothing*) intenta obtener la mejor estimación posible de x_k usando la información, es decir las observaciones, disponibles hasta un instante posterior $k + m$. El filtraje solo dispone de información hasta la iteración k -ésima, y la predicción tiene información hasta una iteración anterior $k - m$. Debido a la cantidad de información que conoce, el suavizado desde luego debería dar el mejor resultado de los tres, seguido por el filtraje y finalmente la predicción.

Si bien en su versión original el filtro de Kalman estaba enfocado en un problema lineal gaussiano [27], han surgido diversas variantes para generalizarlo a otros problemas. Algunas de las más conocidas son el Filtro de Kalman Extendido (EKF) o el Filtro “sin olor” *Uncented* (UKF) [36], que permiten trabajar con sistemas no lineales, el Filtro de Kalman por Ensamblados (EnKF) [37], que es menos costoso computacionalmente y que puede usarse en problemas con gran cantidad de estados, o el Filtro H^∞ que es más robusto ante incertezas en el modelo. Varios de estos filtros tienen versiones continuas y discretas [47].

Esta sección está fuertemente basada en [48] y [49], y se estructura como sigue: en la subsección 1.2.1 comenzaremos estudiando en detalle el filtro de Kalman para el caso más sencillo, en que el sistema es lineal y gaussiano. A continuación, la subsección 1.2.2 presenta el Filtro de Kalman Extendido, una variante que permite trabajar con sistemas no lineales, y que es la que usaremos en el trabajo.

Finalmente se exponen algunas técnicas que serán de utilidad. La subsección 1.2.3 muestra una forma de estimar parámetros desconocidos de la dinámica del sistema. La subsección 1.2.5 muestra una forma de incorporar restricciones a las estimaciones, y la subsección 1.2.4 se aparta del filtraje para presentar una forma de suavizado cuando son conocidas todas las observaciones para un intervalo de tiempo.

1.2.1. Filtro de Kalman lineal discreto

Para comenzar a explicar el filtro de Kalman, se decide utilizar un caso sencillo. Se tiene un sistema como el de la ecuación 1.9, donde $x_k \in \mathbb{R}^n$ y $F_k, G_k \in \mathbb{R}^{n \times n}$ y $w_k \sim \mathcal{N}(0, Q_k)$ es un vector aleatorio que representa un ruido gaussiano en la dinámica del sistema.

$$x_{k+1} = F_k x_k + G_k w_k. \tag{1.9}$$

Los estados x_k del sistema son desconocidos, y solo es posible conocer observaciones y_k de ellos, dadas por la ecuación 1.10, donde $y_k \in \mathbb{R}^m$, $H_k \in \mathbb{R}^{m \times n}$ y $v_k \sim \mathcal{N}(0, R_k)$ representa un ruido gaussiano en las observaciones.

$$y_k = H_k x_k + v_k. \quad (1.10)$$

Para una condición inicial gaussiana $x_0 \sim \mathcal{N}(\hat{x}_0, P_0)$, la linealidad del sistema y el hecho de que el ruido es gaussiano, implica que cada uno de los x_k serán también gaussianos. Se busca estimar x_k a partir de las observaciones y_k . Más específicamente, en términos bayesianos, se busca la distribución *a posteriori* de x_k dado un conjunto de observaciones $y_{1:l} := \{y_1, \dots, y_l\}$. Denotaremos a esta distribución $x_{k,l} := x_k | y_{1:l}$.

Existen tres formas de estimación, que dependen del valor de l ; en primer lugar, si $l < k$, se está intentando estimar x_k con información hasta un instante l anterior a k . A esta forma se le llama predicción o *forecasting*. En segundo lugar, si $l > k$, entonces se está estimando x_k con información hasta un instante l posterior a k . Se llamará a esto suavizado o *smoothing*. Finalmente, si $l = k$ se llamará filtraje o *filtering*. Es esta última forma la que nos presenta mayor interés y que facilita la explicación, por lo que es la forma que será utilizada.

Se sigue el siguiente procedimiento. Se comienza suponiendo que la distribución a posteriori de x_{k-1} con observaciones hasta el instante $k-1$ es conocida, $x_{k-1,k-1} \sim \mathcal{N}(\hat{x}_{k-1,k-1}, P_{k-1,k-1})$; se busca la distribución $x_{k,k}$. Esto se logra mediante una iteración del filtro de Kalman, la cual se divide en dos etapas; una de predicción, que usa la dinámica para obtener $x_{k+1,k}$, y una de análisis, que incorpora y_k para obtener $x_{k,k}$.

La propiedad 1.1, cuya demostración puede leerse en [48] y que ocupa el Teorema de Bayes, será útil a la hora de hacer los cálculos.

Proposición 1.1 Sean X, Y vectores aleatorios conjuntamente gaussianos, ie. tal que

$$Z := \begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}, \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \right).$$

Entonces

$$(X|Y = y) \sim \mathcal{N}(\bar{x} + \Sigma_{xy} \Sigma_{yy}^{-1} (y - \bar{y}), \Sigma_{xx} - \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx}).$$

1. En el paso de predicción se hace una suposición *a priori* de dónde estará el estado real en la próxima iteración a partir de la dinámica y de la estimación actual, sin utilizar la observación y_k . Usando que $x_{k-1,k-1} \sim \mathcal{N}(\hat{x}_{k-1,k-1}, P_{k-1,k-1})$, y la ecuación 1.9, no es difícil ver que $x_{k,k-1} \sim \mathcal{N}(\hat{x}_{k,k-1}, P_{k,k-1})$, con

$$\begin{aligned} \hat{x}_{k,k-1} &= F_{k-1} \hat{x}_{k-1,k-1} \\ P_{k,k-1} &= F_{k-1} P_{k-1,k-1} F_{k-1}^\top + G_k Q_k G_k^\top. \end{aligned} \quad (1.11)$$

2. En el paso de análisis, se incorpora la observación y_k para obtener una estimación *a posteriori*. Para esto basta con usar la proposición 1.1, calculando la distribución

conjunta de x_k e y_k dado $y_{1:k}$, para obtener la expresión 1.12. La propiedad 1.1 nos permite decir que $x_{k,k} \sim \mathcal{N}(\hat{x}_{k,k}, P_{k,k})$, con $\hat{x}_{k,k}$ y $P_{k,k}$ definidos en las ecuaciones 1.13.

$$\begin{pmatrix} x_k \\ y_k \end{pmatrix} \Big| y_{1:k} \sim \mathcal{N} \left(\begin{pmatrix} \hat{x}_{k,k-1} \\ H_k \hat{x}_{k,k-1} \end{pmatrix}, \begin{pmatrix} P_{k,k-1} & P_{k,k-1} H_k^T \\ H_k P_{k,k-1} & H_k P_{k,k-1} H_k^T + R_k \end{pmatrix} \right), \quad (1.12)$$

$$\begin{aligned} K_k &= P_{k,k-1} H_k^T (H_k P_{k,k-1} H_k^T + R_k)^{-1} \\ \hat{x}_{k,k} &= \hat{x}_{k,k-1} + K_k (y_k - H_k \hat{x}_{k,k-1}) \\ P_{k,k} &= P_{k,k-1} - K_k H_k P_{k,k-1}. \end{aligned} \quad (1.13)$$

Estos dos pasos de predicción y análisis entregan una regla que permite iterar; basta con definir la condición inicial como $\hat{x}_{0,-1} := x_0$, $P_{0,-1} := P_0$. El Teorema 1.2, cuya demostración está en [48], asegura que los valores $\hat{x}_{k,k}$ son estimadores de mínima varianza para $x_{k,k}$.

Teorema 1.2 Sean X, Y vectores aleatorios conjuntamente gaussianos. Sea también $\hat{x} := \mathbb{E}(X|Y = y)$. Entonces \hat{x} es un estimador de mínima varianza, es decir, satisface

$$\mathbb{E}(\|X - \hat{x}\|^2 | Y = y) \leq \mathbb{E}(\|X - z\|^2 | Y = y) \quad \forall z$$

Filtro de Kalman Lineal Discreto

Para el caso lineal discreto como el presentado por las ecuaciones 1.14, el filtro de Kalman se define como 1.15, donde $\hat{x}_{0,-1} := x_0$ y $P_{0,-1} := P_0$. Los valores $\hat{x}_{k,k}$ son estimadores de mínima varianza de $x_{k,k}$.

$$\begin{aligned} x_{k+1} &= M_k x_k + F_k w_k \\ y_k &= H_k x_k + v_k. \end{aligned} \quad (1.14)$$

$$\begin{aligned} \hat{x}_{k,k-1} &= F_{k-1} \hat{x}_{k-1,k-1} \\ P_{k,k-1} &= F_{k-1} P_{k-1,k-1} F_{k-1}^T + G_{k-1} Q_{k-1} G_{k-1}^T \\ K_k &= P_{k,k-1} H_k^T (H_k P_{k,k-1} H_k^T + R_k)^{-1} \\ \hat{x}_{k,k} &= \hat{x}_{k,k-1} + K_k (y_k - H_k \hat{x}_{k,k-1}) \\ P_{k,k} &= P_{k,k-1} - K_k H_k P_{k,k-1}. \end{aligned} \quad (1.15)$$

1.2.2. Filtro de Kalman Extendido (EKF)

El filtro de Kalman Extendido permite trabajar con sistemas no lineales. Se trabaja ahora con un modelo como el dado por las ecuaciones 1.16, donde las funciones f_k, g_k, h_k son posiblemente no lineales.

$$\begin{aligned} x_{k+1} &= f_k(x_k, t_k) + g_k(x_k, t_k) w_k \\ y_k &= h(x_k, t_k) + v_k. \end{aligned} \quad (1.16)$$

Usando la notación de las ecuaciones 1.17, es posible definir el filtro de Kalman Extendido como en 1.18. A diferencia del caso lineal, este ya no es óptimo.

$$F_k := \left. \frac{\partial f_k(x, t)}{\partial x} \right|_{\substack{x=\hat{x}_{k,k} \\ t=t_k}} \quad H_k := \left. \frac{\partial h_k(x, t)}{\partial x} \right|_{\substack{x=\hat{x}_{k,k} \\ t=t_k}} \quad G_k := g_k(\hat{x}_{k,k}, t_k). \quad (1.17)$$

Filtro de Kalman Extendido (EKF)

Para el caso lineal discreto como el presentado por las ecuaciones 1.16, el filtro de Kalman Extendido se define como 1.18, donde $\hat{x}_{0,-1} := x_0$ y $P_{0,-1} := P_0$.

$$\begin{aligned} \hat{x}_{k,k-1} &= f_{k-1}(\hat{x}_{k-1,k-1}, t_{k-1}) \\ P_{k,k-1} &= F_{k-1}P_{k-1,k-1}F_{k-1}^\top + G_{k-1}Q_{k-1}G_{k-1}^\top \\ K_k &= P_{k,k-1}H_k^\top (H_kP_{k,k-1}H_k^\top + R_k)^{-1} \\ \hat{x}_{k,k} &= \hat{x}_{k,k-1} + K_k(y_k - h_k(\hat{x}_{k,k-1}, t_k)) \\ P_{k,k} &= P_{k,k-1} - K_kH_kP_{k,k-1}. \end{aligned} \quad (1.18)$$

1.2.3. Estimación de parámetros: estado aumentado

Hasta ahora hemos trabajado con el supuesto de que las únicas cantidades desconocidas eran los vectores x_k . Sin embargo, la mayoría de las veces tendremos además parámetros desconocidos p , los cuales también necesitan ser estimados. Este apartado está basado en la sección 13.4 de [49].

Se tiene un sistema donde las matrices dependen de forma no lineal de cierto parámetro desconocido p , como se ve en las ecuaciones 1.19.

$$\begin{aligned} x_{k+1} &= F_k(p)(x_k) + G_k(p)w_k \\ y_k &= H_kx_k + v_k. \end{aligned} \quad (1.19)$$

Por simplicidad no se agregó dependencia de p a las mediciones, pero es un caso que puede extenderse de manera sencilla a partir de este. Para estimar p se define un estado aumentado x' como en 1.20. Bajo el supuesto de que p es constante, se usa la dinámica $p_{k+1} = p_k + w_{pn}$. El ruido w_{pn} artificial permitirá al filtro cambiar su estimación de p_k . Finalmente, el sistema aumentado está dado por las ecuaciones 1.21.

$$x'_k = \begin{bmatrix} x_k \\ p_k \end{bmatrix}. \quad (1.20)$$

$$\begin{aligned}
x'_{k+1} &= \begin{bmatrix} F_k(p_k)(x_k) + G_k(p_k)w_k \\ p_k + w_{pn} \end{bmatrix} \\
&= f(x'_k, w_k, w_{kp}) \\
y_k &= \begin{bmatrix} H_k & 0 \end{bmatrix} \begin{bmatrix} x_k \\ p_k \end{bmatrix} + v_k.
\end{aligned} \tag{1.21}$$

$f(x'_k, w_k, w_{kp})$ es una función no lineal en el estado aumentado x'_k , por lo que se puede estimar x'_k a partir de algún filtro no lineal como EKF.

1.2.4. Suavizado: *RTS smoother*

Recordamos que $\hat{x}_{k,l}$ corresponde a la estimación de x_k utilizando las observaciones hasta tiempo l , que son denotadas $y|_{1:l}$. Hasta ahora sólo se ha abordado el proceso de filtraje, por lo que solo se ha calculado $\hat{x}_{k,k}$ o $\hat{x}_{k,k-1}$; la estimación del estado x_k solo usa las observaciones hasta el estado k , $y|_{1:k}$, o hasta el estado $k-1$, $y|_{1:k-1}$. En ciertas situaciones, sin embargo, es interesante mejorar una estimación incorporando observaciones obtenidas posteriormente, o de manera más técnica, calcular $\hat{x}_{k,k+1}, \hat{x}_{k,k+2}, \hat{x}_{k,k+3}, \dots$. El proceso de obtener estas estimaciones usando datos de tiempos posteriores es llamado *smoothing* o suavizado. El nombre se debe a que, al tener más información disponible, es posible generar estimaciones con menos ruido, más “suaves”.

Hay tres escenarios comunes donde surge esta necesidad. El primero de ellos es el suavizado de punto fijo, donde se busca estimar un estado en un tiempo fijo, digamos j , y el número de observaciones disponibles aumenta continuamente. Un ejemplo de esto podría ser. En este caso se intenta estimar $\hat{x}_{j,j+1}, \hat{x}_{j,j+2}, \hat{x}_{j,j+3}, \dots$. Un segundo escenario es el suavizado con retardo fijo. En este caso, para cada estado k se tienen N observaciones posteriores, donde N es un número fijo. Se intenta estimar $\hat{x}_{k,k+N}$ para $k = 1, 2, \dots$.

El último escenario es el que será utilizado en este trabajo, el suavizado con intervalo fijo. En este caso, se cuenta con un intervalo de M mediciones $y|_{1:M}$, y se busca la mejor estimación posible de todos los estados hasta tiempo M , es decir, se quiere estimar $\hat{x}_{0,M}, \hat{x}_{1,M}, \hat{x}_{2,M}, \dots, \hat{x}_{M,M}$.

El suavizado de intervalo fijo suele realizarse en dos pasadas: la primera permite obtener estimaciones usando los datos pasados, y la segunda suaviza los resultados. Una de las formas más comunas es la de Rauch, Tung y Striebel [50], conocida como *RTS smoother*. Más detalles en el capítulo 9, *Optimal Smoothing*, de [49]. Para el sistema dado por 1.22, el *RTS smoother* hace una pasada “hacia adelante” o *forward* mediante las ecuaciones 1.24 y luego una “hacia atrás” o *backward* usando 1.26.

$$\begin{aligned}
x_k &= F_{k-1}x_{k-1} + w_{k-1} \\
y_k &= H_k x_k + v_k \\
w_k &\sim \mathcal{N}(0, Q_k) \\
v_k &\sim \mathcal{N}(0, R_k).
\end{aligned} \tag{1.22}$$

1. Inicializar el filtro *forward*.

$$\begin{aligned}\hat{x}_{f0} &= \mathbb{E}(x_0) \\ P_{f0}^+ &= \mathbb{E}[(x_0 - \hat{x}_{f0})(x_0 - \hat{x}_{f0})^\top].\end{aligned}\tag{1.23}$$

2. Para $k = 1, \dots, N$ (donde N es el tiempo final), ejecutar el filtro *forward*, que es simplemente el filtro de Kalman estándar.

$$\begin{aligned}P_{fk}^- &= F_{k-1}P_{f,k-1}^+F_{k-1}^\top + Q_{k-1} \\ K_{fk} &= P_{fk}^-H_k^\top(H_kP_{fk}^-H_k^\top + R_k)^{-1} \\ \hat{x}_{fk}^- &= F_{k-1}\hat{x}_{f,k-1}^+ \\ \hat{x}_{fk}^+ &= \hat{x}_{fk}^- + K_{fk}(y_k - H_k\hat{x}_{fk}^-) \\ P_{fk}^+ &= (I - K_{fk}H_k)P_{fk}^-(I - K_{fk}H_k)^\top + K_{fk}R_kK_{fk}^\top \\ &= [(P_{fk}^-)^{-1} + H_k^\top R_k^{-1}H_k]^{-1} \\ &= (I - K_{fk}H_k)P_{fk}^-.\end{aligned}\tag{1.24}$$

3. Inicializar el *RTS smoother*

$$\begin{aligned}\hat{x}_k &= \hat{x}_{fN}^+ \\ P_k &= P_{fN}^+.\end{aligned}\tag{1.25}$$

4. Para $k = N - 1, \dots, 1, 0$ ejecutar las ecuaciones del *RTS smoother*

$$\begin{aligned}\mathcal{I}_{f,k+1}^- &= (P_{f,k+1}^-)^{-1} \\ K_k &= P_{fk}^+F_k^\top\mathcal{I}_{f,k+1}^- \\ P_k &= P_{fN}^+ - K_k(P_{f,k+1}^- - P_{k+1})K_k^\top \\ \hat{x}_k &= \hat{x}_{fk}^+ + K_k(\hat{x}_{k+1} - \hat{x}_{f,k+1}^-).\end{aligned}\tag{1.26}$$

1.2.5. Restricciones: observaciones (casi) perfectas

El survey [51] presenta varias alternativas para tratar con restricciones. La idea de la técnica de observaciones (casi) perfectas es la siguiente: en un sistema de la forma $x_{k+1} = f_k(x_k) + w_k$, observado mediante una ecuación $y_k = Hx_k + v_k$, se busca imponer una restricción $|Dx - d| \leq \varepsilon$. Esto se consigue ampliando la matriz de observaciones como muestra la ecuación 1.27. Si se busca una restricción fuerte con $\varepsilon = 0$, la técnica es llamada “observaciones perfectas”, y si se busca una restricción suave con $\varepsilon > 0$, recibe el nombre de “observaciones casi perfectas”.

$$\begin{bmatrix} y_k \\ d \end{bmatrix} = \begin{bmatrix} H \\ D \end{bmatrix} x_k + \begin{bmatrix} v_k \\ \varepsilon \end{bmatrix}.\tag{1.27}$$

1.3. Antecedentes: El caso del COVID-19

El COVID-19 se transmite por microgotas emitidas al respirar, hablar, toser o estornudar, o por contacto estrecho [2][3]. Se ha intentado mitigar mediante diferentes medidas [4][5], que incluyen varias formas de distanciamiento social como reducciones de movilidad voluntarias, cuarentenas totales o parciales, teletrabajo, distancias mínimas entre personas, etc. Se han fomentado además distintas medidas de higiene como el uso de mascarillas, lavado frecuente de manos, ventilación de espacios, entre otras.

Con respecto a las medidas farmacéuticas, durante 2020 se estudiaron y desarrollaron varios intentos de vacunas, utilizando conocimientos adquiridos en la lucha contra virus como SARS-CoV y MERS-CoV. En 2021 se comenzó la vacunación masiva y hacia marzo de 2022, más de la mitad de la población mundial ha recibido al menos una dosis, con países como China y Singapur teniendo más de 85% de su población vacunada según datos publicados en *Our World in Data* [6].

Con respecto a características particulares de la enfermedad, se conoce de la existencia de casos sin síntomas o con síntomas leves y que también transmiten el virus, especialmente si no son reportados [52][53][54]. Su tiempo de incubación ha sido estimado en poco más de 5 días (5.1 días, Intervalo de Confianza del 95% de 4.5 a 5.8 días según [55]; 5.2 días, IC 95% de 4.1 a 7.0 días según [56]).

El impacto de la enfermedad ha sido heterogéneo entre la población. Se sabe que la edad es un factor a considerar; la susceptibilidad a la infección y la probabilidad de desarrollar síntomas luego del contagio aumentan con la edad [7]. Similarmente, tanto el porcentaje de infectados cuya gravedad requiere hospitalización como el porcentaje que termina falleciendo es más alto entre los adultos mayores [8]. El nivel socioeconómico tampoco puede ser ignorado; el menor acceso a la salud [9], sumado al desempleo y la existencia de enfermedades crónicas previas [10] son un agravante que hace más vulnerables a los niveles socioeconómicos más pobres.

Los datos de movilidad han sido utilizados ampliamente para modelar el avance de la enfermedad [11][12][13]; en la mayoría de los países la movilidad explica una parte importante de las variaciones en la transmisibilidad [14]. El artículo [15] nota como las diferencias en movilidad explican las diferencias en transmisión en diferentes grupos económicos, notando que además las clases de menor nivel socioeconómico se enfrentan a ambientes más riesgosos.

1.3.1. El brote de COVID-19 en Santiago de Chile

La Región Metropolitana (RM) contiene a la ciudad de Santiago, la capital de Chile y su centro administrativo. Está formada por 52 comunas, las cuales varían significativamente en términos socioeconómicos. Su población es de unos 7.3 millones de personas, un 40% aproximadamente de la población del país.

Las primeras medidas tomadas por el Gobierno de Chile para enfrentar la pandemia se centraron en fortalecer la capacidad hospitalaria: aumento en la cantidad de camas UCI, in-

tegración de los sistemas de salud público y privado, fortalecimiento de la capacidad de testeo y trazabilidad, preparación de residencias sanitarias, importación de material y equipamiento sanitario.

Se implementaron además [57] medidas de distanciamiento social como la prohibición de eventos de gran concurrencia y el cierre de instituciones educativas. Se llevaron a cabo cuarentenas localizadas, hasta entrar en cuarentena total a mediados de mayo de 2020. En julio, una vez superado el primer *peak* de casos, se comienza a aplicar el Plan Paso a Paso, una estrategia de desconfiamiento gradual que sigue operando con modificaciones hasta la fecha.

Ya a fines de mayo de 2020, [31] había hecho notar "la dificultad de implementar cuarentenas obligatorias en comunas de menores recursos, donde la gente vive al día y genera sus ingresos con su trabajo diario", sugiriendo "que al aplicar las cuarentenas en comunas de menores niveles socioeconómicos, el gobierno las complemente con herramientas que apoyen a sus habitantes y provea de recursos para cubrir necesidades urgentes". Algunas medidas tomadas en esta dirección fueron la entrega de bonos e ingresos suplementarios, beneficios para el pago de servicios básicos, canastas de alimentos, etc.

Sin embargo, en [33] se muestra que, si bien las cuarentenas fueron efectivas en las comunas de mayores ingresos, no parecen tener un efecto significativo en las comunas de menores ingresos. Atribuye esto en parte a las diferencias en el cumplimiento de las medidas en términos de movilidad, así como a las diferentes capacidades de testeo. Estos resultados son similares a los obtenidos en [34]. En [32] se explora aún más este tópico, encontrando además una fuerte correlación entre el nivel socioeconómico y la mortalidad, que afecta aún más a los menores de 40 años.

El 3 de febrero de 2021 comienza el proceso de vacunación masiva de la población [58], el cual se realiza en su mayor parte con las vacunas Sinovac, Pfizer-BioNTech y Oxford-AstraZeneca. Hacia marzo de 2022 se está aplicando la cuarta dosis de refuerzo; Chile se encuentra entre los países con mayor porcentaje de personas vacunadas, rondando el 90% de la población inoculada con dos dosis, similar a países como Canadá y Singapur [6].

El Ministerio de Salud ha liberado varios conjuntos de datos, que incluyen series de tiempo de casos confirmados, UCI, fallecidos, varias de ellas agregadas por edad o comuna. Están disponibles en el GitHub del Ministerio de Ciencia [35].

Capítulo 2

Metodología

Una vez presentadas las bases teóricas necesarias para entender el trabajo, este capítulo describe los procedimientos llevados a cabo para cumplir los objetivos. Recordamos que el objetivo principal del trabajo es estimar el factor sanitario de distintas clases sociales, por medio del *framework* lagrangiano de clases y ambientes de [18]. Para esto, en primer lugar, se necesita definir una metodología que permita estimar el factor sanitario, y en segundo lugar, se debe implementar y evaluar la metodología, aplicándola al caso de estudio: la pandemia de COVID-19 en Santiago.

El trabajo realizado se divide en varias etapas. En primer lugar, es necesario explorar las distintas fuentes de datos disponibles, tanto de movilidad y viajes como las series de tiempo epidemiológicas, como cantidad de casos registrados, fallecidos, etc. Esto se reporta en la sección 2.1. Esto permite decidir el modelo específico a utilizar, lo que incluye la elección de compartimientos a utilizar (susceptibles, expuestos, infectados, etc), parámetros, además de las clases y ambientes. Todo esto debe considerar las características del COVID-19 y también los datos disponibles. Todo esto se hace en la sección 2.2.

Una vez decididas las clases y ambientes, es necesario estimar una matriz de tiempos de residencia que diga cuánto tiempo pasa cada clase en cada ambiente. Se desea que esta matriz sea variable en el tiempo, de forma que incorpore cambios causados por las medidas de mitigación como cuarentenas, teletrabajo, etc. Esto se presenta en la sección 2.3.

Todo lo anterior ya permite correr un modelo de la ciudad de Santiago, definiendo factores sanitarios sintéticos y obtener simulaciones. Se busca ahora estimar el factor sanitario a partir de observaciones del modelo. Para esto se utiliza Filtro de Kalman, y es necesario elegir e implementar una de sus muchas variantes (extendido, *unscented*, filtro por ensambles, etc), y elegir cómo usar el filtro para hacer estimación de parámetros (estado aumentado, filtro de Kalman con múltiples modelos, etc). Posteriormente se estiman los factores sanitarios para cada una de las clases elegidas, ajustando el modelo a los datos reales de la ciudad de Santiago. Para esto es necesario definir qué variables del modelo observar, que datos utilizar como observación y cómo ajustar los parámetros específicos del filtro. Todo esto se encuentra en la sección 2.4.

Finalmente, se utiliza la matriz de tiempos de residencia y los factores sanitarios obtenidos

para generar varios casos hipotéticos; distintas variaciones en cuarentenas, diferentes niveles de cuidado. Esto entrega varios desarrollos hipotéticos de la pandemia, y se usará para evaluar el modelo. Los casos a utilizar se exponen en la sección 2.5.

2.1. Datos disponibles

Esta sección describe las distintas fuentes de datos consideradas en el estudio. Los datos de interés pueden ser clasificados en tres categorías. En primer lugar, se busca caracterizar a la población de la Región Metropolitana, en términos etarios y socioeconómicos. Esto se logra mediante el Censo 2017 y el Índice de Prioridad Social 2019, en las subsecciones 2.1.1 y 2.1.2 respectivamente.

En segundo lugar, se quiere conocer el comportamiento de la población, específicamente su uso del tiempo. Para esto se hace uso de la relación entre viajes y tiempos de actividad; existen modelos [22][23] de predicción de demanda de viajes que utilizan un enfoque basado en actividades, y [24] obtiene tiempos de actividad a partir del registro de viajes de Encuestas Origen-Destino. En esa dirección, se considera la Encuesta Origen Destino Santiago 2012, descrita en la subsección 2.1.3, ya que es una fuente bastante rica; provee información muy granular de personas, viajes y sus propósitos. Datos más actualizados que capturan las variaciones debido a la pandemia son los Informes de Movilidad Local sobre COVID-19, presentados en la subsección 2.1.4, y el Uso de infraestructura de telecomunicaciones, en la subsección 2.1.5.

Finalmente, se requieren datos para entender el avance de la pandemia en la Región Metropolitana. Datos-COVID19, un repositorio [35] mantenido por la Mesa de Datos del Ministerio de Ciencia, Tecnología, Conocimiento e Innovación, contiene distintas series de tiempo de casos confirmados, hospitalizados, fallecidos, vacunados en el país, con distintos niveles de agregación. Se comenta en más detalle en la subsección 2.1.6.

2.1.1. Censo 2017

El Censo de Población y Vivienda 2017 fue un proceso liderado por el Instituto Nacional de Estadística (INE), y permitió contar y caracterizar a 17.574.003 personas y 6.499.355 viviendas en todo el territorio nacional.

Para el levantamiento del Censo del año 2017 y la desagregación de los datos censales se utiliza en primer lugar la División Política Administrativa del país; regiones y comunas. Luego, el territorio de cada comuna se divide en distritos censales, los que pueden ser urbanos, rurales o mixtos. A su vez, en el área urbana se reconocen zonas censales, las que están compuestas de manzanas, y en el área rural, localidades.

El Censo está disponible en <http://www.censo2017.cl/>. Los datos de interés son la población de la Región Metropolitana, desagregada por zona censal o comuna, sexo y edad.

2.1.2. Índice de prioridad social (IPS) 2019

La Secretaría Regional Ministerial de Desarrollo Social y Familia de la Región Metropolitana de Santiago propone en [1] una metodología que permite comparar las comunas respecto de sus niveles de desarrollo socioeconómico. El Índice de Prioridad Social (IPS) es un indicador compuesto que contempla ingresos, educación y salud. Se trata de un índice sintético cuyo valor numérico permite dimensionar el nivel de vida alcanzado por la población de una comuna, relativo al de las demás.

Cada comuna de la Región Metropolitana tiene un IPS asociado, el cual toma valores de 0 a 100, siendo 100 la prioridad máxima o mayor vulnerabilidad, y 0 sin prioridad. Las comunas pueden ser clasificadas de acuerdo a su IPS; [1] ofrece una clasificación en cinco categorías de prioridad social:

- **Alta:** incluye comunas como La Pintana y Cerro Navia.
- **Media Alta:** comunas como San Bernardo y El Bosque.
- **Media Baja:** Quinta Normal y La Granja.
- **Baja:** como Santiago y San Miguel.
- **Sin Prioridad:** como Las Condes y Providencia.

2.1.3. Encuesta Origen-Destino Santiago 2012

Las Encuestas de Movilidad constituyen la principal fuente de información utilizada en todo proceso de planificación de los sistemas de transporte. Éstas entregan antecedentes relevantes sobre los patrones de movilidad de una determinada ciudad y proporcionan los datos requeridos para la calibración de los modelos de análisis de transporte. Estas encuestas son realizadas por el Programa de Vialidad y Transporte Urbano SECTRA, de la Subsecretaría de Transportes del Ministerio de Transportes y Telecomunicaciones (MTT) y se realizan cada diez años en las ciudades más grandes del país. La encuesta más reciente de la ciudad de Santiago disponible al momento de realizar el trabajo es la Encuesta Origen-Destino (EOD) Santiago 2012 [59]; en enero de 2021 se discutía [60] una metodología basada en *Big Data* para la nueva versión de la EOD, pero los resultados no han sido publicados aún. Los datos, por tanto, pueden estar desactualizados.

La EOD Santiago 2012 recopiló información de los residentes de 18.000 hogares de Santiago, seleccionados aleatoriamente, durante el periodo comprendido entre julio de 2012 y noviembre de 2013. Su objetivo fue conocer las características de los viajes que se realizan en la ciudad y de quienes los efectúan. La encuesta se realizó mediante entrevista personal y los datos fueron recolectados en días hábiles y fines de semana, tanto en temporada normal como estival. En total se encuestaron alrededor de 60.000 personas.

Como resultado, se cuenta con una base de datos con información de cada hogar, de las personas encuestadas y de los viajes realizados. Para cada hogar se cuenta con la fecha y

temporada en que se realizó la encuesta, la ubicación geográfica, número de personas que lo habitan e ingresos. De cada persona se conoce el hogar al que pertenece y la relación que tiene con el resto de moradores, sexo, edad, ocupación y tramo de ingresos. De cada viaje se conoce la persona que lo realizó, la zona censal de origen y destino, la hora de inicio y fin del viaje, los medios de transporte usados y el propósito del viaje. Se conocen todos los viajes realizados por cada persona. Estos datos se encuentran disponibles en http://www.sectra.gob.cl/encuestas_movilidad/encuestas_movilidad.htm.

2.1.4. Informes de Movilidad Local sobre COVID-19

Son reportados por Google a partir de datos anonimizados que usan en servicios como Google Maps. Pretenden proporcionar información a las autoridades sanitarias de los cambios en la movilidad como consecuencia de las distintas políticas de contención del COVID-19. Las series de tiempo se encuentran clasificadas en varias categorías: tiendas y ocio, supermercados y farmacias, parques, estaciones de transporte, lugares de trabajo y zonas residenciales. Los datos pueden verse en la figura 2.1, y se encuentran disponibles en <https://www.google.com/covid19/mobility/index.html?hl=es>.

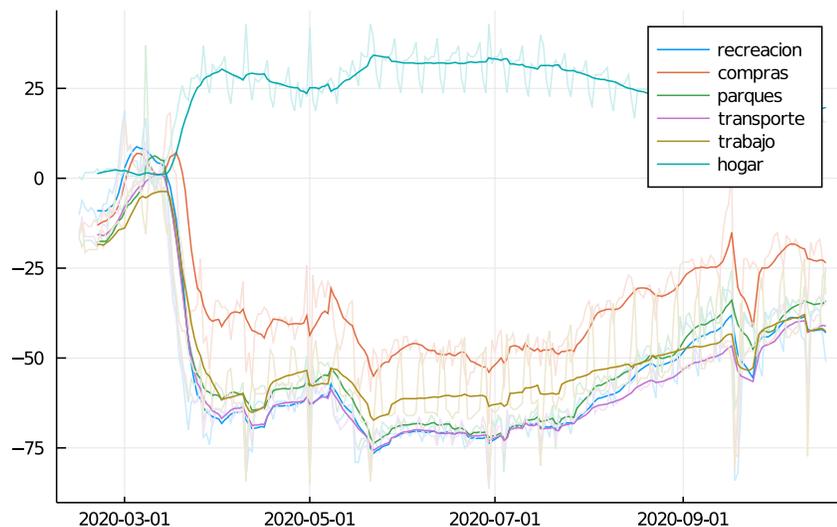


Figura 2.1: Variación en la movilidad agregada por actividad en la Región Metropolitana y media móvil de 7 días. Fuente: Elaboración propia con datos de Google.

2.1.5. Uso de infraestructura de telecomunicaciones

El Instituto Sistemas Complejos de Ingeniería (ISCI), que agrupa a un conjunto de investigadores de varias universidades de Chile con el objeto de generar trabajo científico y desarrollar soluciones para problemas complejos de ingeniería, en conjunto con Entel Ocean, la Unidad Digital de Entel, compañía de tecnología y telecomunicaciones, recopilaron información acerca del uso de infraestructura de telecomunicaciones.

Los datos se encuentran agrupados a nivel de zona censal. La información disponible les permitió deducir la zona hogar, en donde las personas se encuentran frecuentemente en horarios no laborales. Para cada día laboral (lunes a viernes), determinaron el flujo desde cada zona hogar a otras zonas, durante horarios de trabajo. Estos flujos pueden ser dentro de la misma comuna o a otras comunas, y fueron considerados un proxy de la movilidad en Santiago.

Las dos primeras semanas de marzo del 2020, antes de la declaración de la fase 2 de pandemia por COVID-19 en Chile, fueron consideradas como semanas “base”, siendo una aproximación para la movilidad usual de cada zona [31]. Los cambios en movilidad por zona con respecto a esa movilidad base fueron reportados para cada semana, y se encuentran disponibles en el repositorio de la Mesa de Datos del Ministerio de Ciencia, Tecnología, Conocimiento e Innovación [35]. También pueden verse en el Visor Movilidad en <https://covidanalytics.isci.cl/movilidad/visor-movilidad/>, como en la figura 2.2.

La versión completa de estos datos, que contiene información tanto del origen como del destino de los transeúntes, fue analizada en [61]. La versión pública de estos datos, sin embargo, solo contiene la movilidad de entrada y salida de cada zona censal/comuna; es posible saber cuánta gente sale y entra a una zona censal o comuna pero no hacia dónde va.

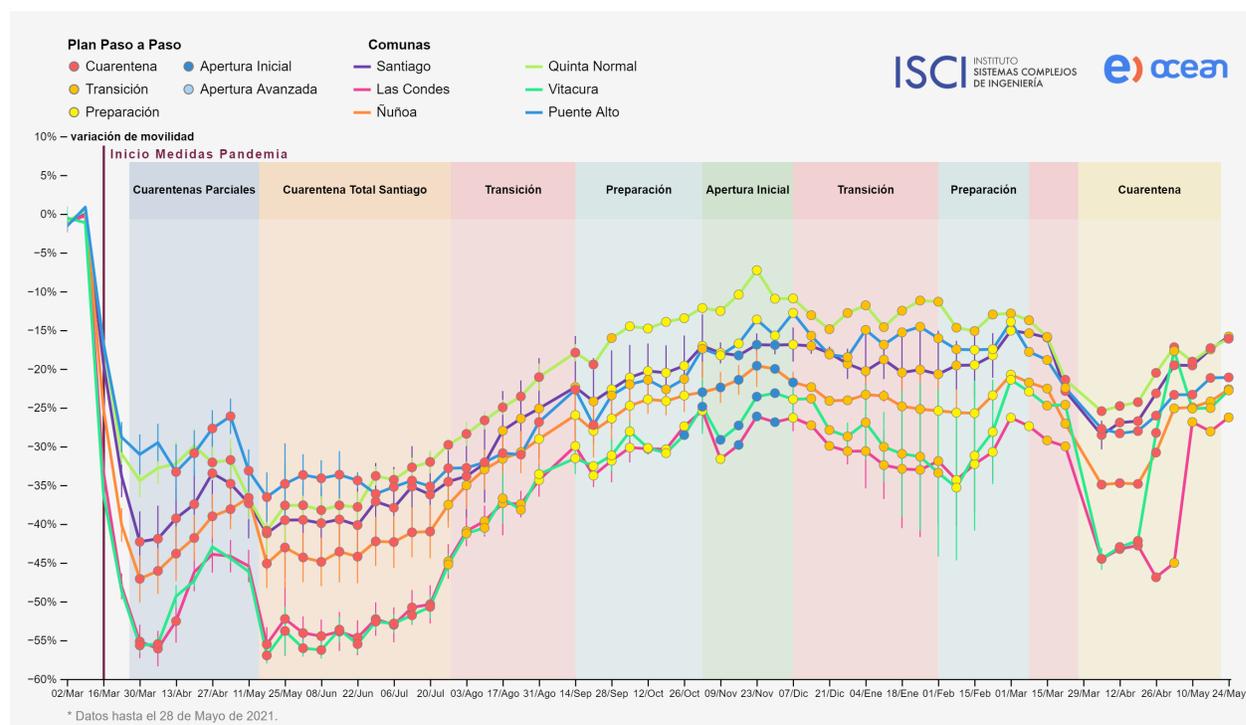


Figura 2.2: Variación en la movilidad de seis comunas de la Región Metropolitana. Fuente: ISCI Covid Analytics.

2.1.6. Datos-COVID19

La Mesa de Datos COVID-19 liderada por el Ministerio de Ciencia, Tecnología, Conocimiento e Innovación busca proveer de información del país durante la pandemia, con el fin de

promover el uso de datos para investigación científica, clínica y para soluciones innovadoras. Dispone de datos documentados del Ministerio de Salud (MINSAL) y de otras fuentes [35].

Fallecidos a nivel nacional

Cantidad total de casos fallecidos acumulados

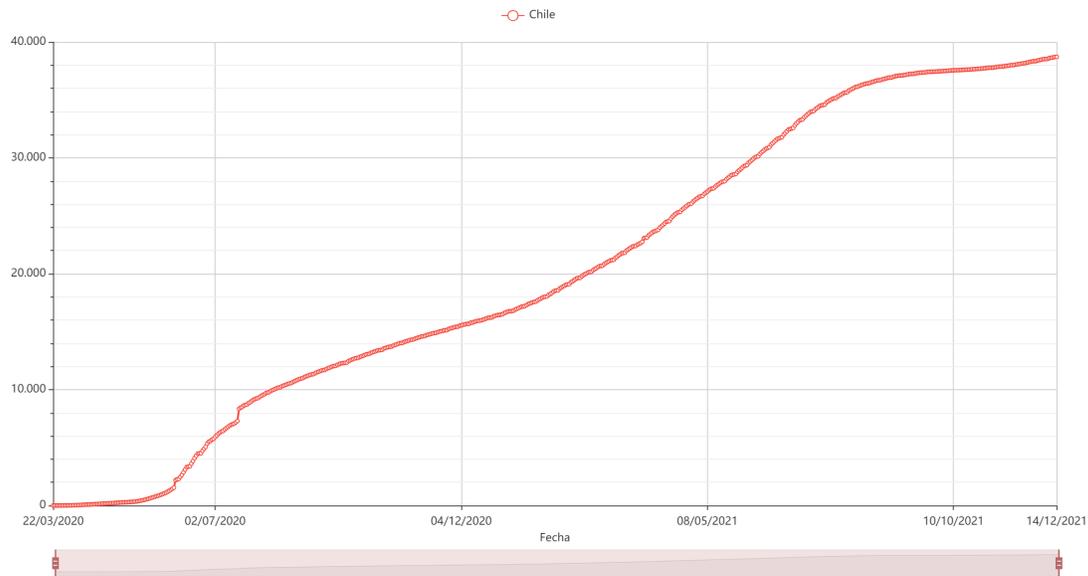


Figura 2.3: Total de casos fallecidos por COVID-19 acumulados a nivel nacional. Fuente: Centro de Modelamiento Matemático con Datos-COVID19.

Cobertura Vacunación a nivel regional

Cobertura de vacunación Covid-19, con respecto a la población total

Use el botón 'Otras regiones' en la barra superior para cambiar la(s) región(es) a visualizar

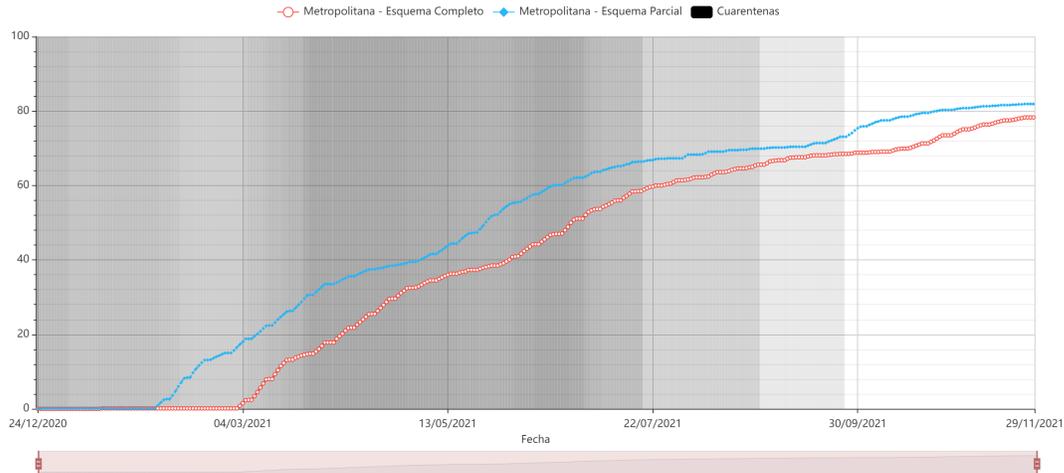


Figura 2.4: Porcentaje de cobertura vacunación COVID-19 en la Región Metropolitana. Fuente: Centro de Modelamiento Matemático con Datos-COVID19.

La información está organizada en distintos *Data Products*. Existen series de tiempo de casos confirmados, hospitalizados, UCI, conectados a ventilación mecánica, fallecidos, cantidad de exámenes PCR realizados, nacimientos y fallecimientos, cuarentenas, etc. Estos datos son utilizados, por ejemplo, en el visualizador del Centro de Modelamiento Matemático

disponible en <https://covid-19vis.cmm.uchile.cl/geo>. Algunos ejemplos de estos datos son las figuras 2.3 y 2.4.

2.2. Decisiones específicas acerca del modelo

En esta sección se tomarán decisiones específicas concernientes a la aplicación de la metodología al caso de estudio. Esto permite decidir el modelo específico a utilizar, lo que incluye la elección de compartimientos a utilizar (susceptibles, expuestos, infectados, etc), parámetros, además de las clases y ambientes. Todo esto debe considerar las características del COVID-19 y también los datos disponibles.

Ya se ha expuesto previamente en 1.1.3 el modelo de dispersión virtual presentado por [17], el cual permite trabajar simultáneamente con clases y ambientes. Ahora se discute las particularidades del modelo a usar en este caso específico. Es necesario definir los compartimientos y parámetros específicos a usar, considerando la enfermedad a modelar, en este caso el COVID-19.

Dependiendo del aspecto de interés, distintos compartimientos han sido usados para modelar el COVID-19. Los modelos revisados por la revisión [62] utilizan los clásicos Susceptibles, Infectados, Recuperados, en combinación con Expuestos, Hospitalizados y/o Fallecidos. La existencia de casos asintomáticos que portan el virus y pueden contagiarlo pero no presentan síntomas (o presentan síntomas muy leves) es considerado en algunos modelos. Algunos consideran compartimientos En Cuarentena y el desarrollo de vacunas da lugar a modelos con compartimientos de Vacunados.

Como se exponía inicialmente, se buscaba un modelo epidemiológico que considere un riesgo en tres factores: amenaza, exposición y vulnerabilidad. Suponiendo que hay varias clases $i \in 1 \dots n$ y ambientes $j \in 1 \dots m$, la idea propuesta de amenaza era de la forma $\beta_j I_j / N_j$, que incluía la fracción de infectados en un lugar y un factor β_j dependiente de las características del lugar. Se consideraban además valores p_{ij} , que representaban la fracción de tiempo que la clase i pasaba en el ambiente j , y un factor sanitario α_i dependiente de la clase i y relacionado a las medidas de cuidado personal y la susceptibilidad al contagio. Al combinar lo anterior se obtenía la expresión 2.1, que da una idea de la tasa de contagios de la clase i buscada.

$$\alpha_i \sum_{j=1}^m \beta_j p_{ij} \frac{I_j}{N_j}. \quad (2.1)$$

Para hacer de esta idea algo concreto, se vuelve al apartado 1.1.3, donde se habló del modelo de clases y ambientes propuesto por [17][18]. Como antes, se separa a la población en n clases, las cuales interactúan en m ambientes o áreas de riesgo. $N_i(t)$, $i = 1, \dots, n$ corresponde a la población de la clase i . Se supone que la población de la clase i pasa una proporción $p_{ij} \in [0, 1]$ de su tiempo en el ambiente j , con $\sum_{j=1}^m p_{ij} = 1$ para cada i . Para casos extremos, por ejemplo, podría darse que $p_{ij} = 0$, es decir, la clase i no gasta nada de su tiempo en el ambiente j , mientras que $p_{ij} = 1$ significa que la clase i pasa todo su tiempo

en el ambiente j . La matriz $P = \{p_{ij}\}_{i=1\dots n, j\dots m}$ es llamada matriz de tiempos de residencia. Esta matriz será estimada a partir de los datos disponibles. Más aún, se busca una matriz que sea variable en tiempo, de tal forma que refleje los cambios en el comportamiento de la población a lo largo de la pandemia.

Se decide trabajar con el ampliamente utilizado modelo SEIR; ya ha sido utilizado para estudiar el impacto de las cuarentenas y reducciones de movilidad y viajes [11][13], además del uso de mascarillas [63], por lo que parece una elección razonable. Las ecuaciones del modelo elegido son 2.2. Se decidió despreocupar los efectos demográficos de los nacimientos y fallecimientos naturales de la población. Eso se traduce en tasas de natalidad y mortalidad b_i, d_i iguales a 0. Esto debido a que no se trabajará a largo plazo, solo con un intervalo de tiempo de poco más de un año.

La matriz de tiempos de residencia $P(t) = \{p_{ij}(t)\}_{i,j}$ será una variable en el tiempo, la forma de estimarla se expondrá en la sección 2.3. Esto se hará de antemano, antes de correr el modelo, por lo que se considera conocida. Se usará la información conocida acerca del COVID-19 para los valores γ_{Ei}, γ_{Ii} y los riesgos específicos de cada ambiente β_j se elegirán de antemano usando valores plausibles. La estimación del factor sanitario $\alpha_i(t)$ se presentará en 2.4.

$$\begin{aligned}
S'_i(t) &= -\lambda_i(\vec{x}, t) \\
E'_i(t) &= \lambda_i(\vec{x}, t)S_i(t) - \gamma_{Ei}E_i(t) \\
I'_i(t) &= \gamma_{Ei}E_i(t) - \gamma_{Ii}I_i(t) \\
R'_i(t) &= \gamma_{Ii}I_i(t) \\
\lambda_i(\vec{x}, t) &= \alpha_i(t) \sum_{j=1}^m \beta_j p_{ij}(t) S_i(t) \left(\frac{\sum_{k=1}^n p_{kj}(t) I_k}{\sum_{k=1}^n p_{kj}(t) N_k} \right).
\end{aligned} \tag{2.2}$$

Con respecto a los compartimientos elegidos, se decidió no utilizar uno de Fallecidos. Ahora bien, [32] menciona que al trabajar con niveles socioeconómicos, la cantidad de fallecidos es más confiable que la de casos infectados detectados debido a problemas de trazabilidad en las comunas de menores recursos, especialmente al inicio de la pandemia. El problema es que la tasa de fallecimiento o la fracción de infectados que fallecen, al igual que el factor sanitario, es probablemente variable en tiempo, dependiente de las distintas políticas usadas para enfrentar la pandemia y además del nivel socioeconómico. Ante la sospecha de que intentar estimar dos parámetros con estas características simultáneamente fuera problemático, se decidió utilizar un modelo más sencillo que solo considera el contagio y no la muerte. También por simplicidad no fue considerada la vacunación, sin embargo, se espera que sus efectos se reflejen directamente en el factor sanitario, al hacer a los vacunados menos vulnerables a una posible infección.

2.3. Matriz de tiempos de residencia

En la sección anterior se planteó el modelo multiclase multiambiente a utilizar. La principal diferencia con respecto a otros modelos multiclase está en la forma en que se calculaba

la tasa de contagio; las n clases interactúan en m ambientes virtuales. Esta interacción está codificada por medio de la matriz de tiempos de residencia $P = \{p_{i,j}\}_{i \in 1 \dots n, j \in 1 \dots m}$, donde la entrada $p_{i,j}$ corresponde a la fracción del día que la clase i pasa en el ambiente j .

El objetivo de este apartado es elegir las clases y ambientes a utilizar para estudiar el caso de Santiago, y estimar la matriz P de tiempos de residencia asociada. Es importante notar que se desea que esta matriz sea variable en el tiempo, puesto que las diferentes medidas de mitigación como cierre de instituciones educativas, cuarentenas, etc. implementadas por el gobierno de Chile han cambiado significativamente el modo de vida en la ciudad.

Se comienza eligiendo las clases y ambientes a utilizar, en la subsección 2.3.1. Posteriormente, en 2.3.2, se describirá la metodología para la obtención de una matriz de tiempos de residencia P^0 correspondiente a la ciudad de Santiago en condiciones normales. Finalmente, en 2.3.3, se describirá la metodología para modificar esta matriz, con el fin de reflejar las variaciones en el comportamiento de la población a lo largo del desarrollo de la pandemia, lo que da lugar a la matriz $P(t)$ definitiva. Esta idea de estimar una matriz de comportamiento normal para luego modificarla ha sido desarrollada también por [11].

2.3.1. Elección de clases y ambientes

El impacto de la enfermedad ha sido heterogéneo entre la población. Se sabe que la edad es un factor a considerar; la susceptibilidad a la infección y la probabilidad de desarrollar síntomas luego del contagio aumentan con la edad [7]. Similarmente, tanto el porcentaje de infectados cuya gravedad requiere hospitalización como el porcentaje que termina falleciendo es más alto entre los adultos mayores [8]. El nivel socioeconómico tampoco puede ser ignorado; el menor acceso a la salud [9], sumado al desempleo y la mayor presencia de enfermedades crónicas previas [10] son un agravante que hace más vulnerables a los niveles socioeconómicos más pobres.

Hay varios criterios a considerar a la hora de elegir las clases y los ambientes. En primer lugar, es ampliamente conocido [7][8] que la edad es un factor importante en el efecto de la enfermedad en una persona; el COVID-19 afecta con más intensidad a los adultos mayores. Varios estudios [9][10][15][32][33] han mostrado que la pandemia ha afectado más fuertemente a los niveles socioeconómicos o comunas más pobres y lo asocian, entre otras cosas, a la dificultad de ellos de acatar las cuarentenas, junto con menor trazabilidad y peor acceso a cuidados de la salud. Puesto que la Encuesta Origen Destino muestra diferencias en el uso del tiempo entre hombres y mujeres [64], el sexo podría ser una variable a considerar también.

Los datos disponibles considerados en 2.1 han de ser considerados también. La Encuesta Origen Destino, por un lado, posee datos muy granulares; es posible conocer para cada individuo su sexo, zona censal y comuna donde reside, edad e ingreso familiar. Más aún, los distintos propósitos de los viajes realizados (regreso al hogar, al trabajo, escuela, compras, transporte público, bicicleta, entre otros) entregan varios ambientes donde ese individuo pasa su tiempo.

Por otro lado, las demás fuentes de datos no poseen tanta información. El Uso de infraestructura de telecomunicaciones permite conocer cómo ha cambiado la movilidad de entrada o

salida a cada zona censal/comuna, pero no distingue entre las actividades que realizan las personas mientras se mueven por la ciudad. Los Informes de Movilidad Local sobre COVID-19 de Google permiten distinguir distintas actividades, pero no permiten separar por comuna. Las series de tiempos de Datos-COVID19 presentan dificultades similares. Están desagregadas por edad, sexo o comuna, pero no por los tres criterios simultáneamente.

Puesto que se desea considerar el efecto de las cuarentenas en el uso del tiempo de las personas a lo largo de la pandemia, sabiendo la importancia que juega en las diferencias entre las distintas comunas, se decide usar los datos de movilidad vistos en la sección 2.1.5. Se decide desechar la información de los distintos ambientes; solo se sabe que hay más o menos movilidad fuera del hogar, pero no se sabe exactamente en qué actividad. El trabajo realizado considera, por tanto, solo $m = 2$ ambientes: **hogar** y **exterior**. Es interesante mencionar que el modelo del *Imperial College* [65] considera que del total de contagios, aproximadamente 1/3 ocurren en el hogar, 1/3 en el trabajo y/o entidades educativas, y el último 1/3 en otras actividades como compras o recreación, en concordancia con matrices de mezcla social como [19].

Los datos elegidos permiten utilizar la información a nivel de zona censal, pero los datos epidemiológicos no son tan granulares, por lo que es más factible distinguir por comuna. Los datos epidemiológicos están disponibles a este nivel también, por lo que es una buena posibilidad. Eso, sin embargo, plantearía un sistema con unas 50 clases, lo que para 4 compartimientos da unas 200 variables; es posible, aunque agrega más carga computacional y dificulta el análisis posterior. Se decide, por simplicidad, reducir el análisis a $n = 5$ categorías socioeconómicas, las mismas propuestas por [1], que surgen tras ordenar las comunas por IPS, como describe la sección 2.1.2. La tabla 2.1 muestra un resumen de las categorías elegidas, la versión completa se encuentra en la tabla 1.

Prioridad Social	IPS mínimo	IPS máximo	Ejemplos
Alta	78.24	83.03	La Pintana, Lo Espejo
Media Alta	73.84	77.39	San Bernardo, El Bosque
Media Baja	65.05	71.36	La Cisterna, Quinta Normal
Baja	53.34	64.37	Santiago, Providencia
Sin Prioridad	6.26	37.36	Las Condes, Lo Barnechea

Tabla 2.1: Resumen de las categorías socioeconómicas por IPS, dadas por [1].

2.3.2. Matriz de Santiago en condiciones normales

En [24] se propone un método para obtener información acerca del uso del tiempo de una persona a partir de los tiempos de inicio y fin de sus viajes y el propósito declarado de este, datos contenidos en la Encuesta Origen-Destino. Para esto supone que el primer viaje comienza desde el hogar y luego cada período de tiempo entre dos viajes consecutivos es asignado a la actividad declarada como propósito en el viaje anterior. La Figura 2.5 es un diagrama de este proceso.

Algoritmo 1: Tiempos de residencia a partir de una lista de viajes

Entrada: Una lista de personas P . Para cada $p \in P$, una lista V_p de todos los viajes hechos por p , ordenados del primero al último. Para un viaje v es posible obtener t_0, t_f tiempos de inicio y fin, modo de transporte y propósito del viaje. Dos funciones $\mathcal{J}_m : \text{modo} \mapsto j$ y $\mathcal{J}_p : \text{propósito} \mapsto j$ que asocian cada modo de transporte (bus, auto, etc) y cada propósito de viaje (al trabajo, de compras, etc) a un ambiente $j \in \{1, \dots, m\}$ (trabajo, compras, transporte público, etc). Un ambiente por defecto j_0 (hogar), donde las personas sin viajes pasan su tiempo, y donde las personas comienzan su primer viaje. Una función $\mathcal{C} : p \in P \mapsto i$ que asigne a cada persona p una clase $i \in \{1, \dots, n\}$. Tiempos de inicio T_0 y fin T_f del intervalo de tiempo total considerado.

Resultado: Una matriz de tiempos de residencia $R = (r_{ij})_{i=1\dots n, j=1\dots m}$ donde r_{ij} es la fracción del tiempo que la clase i pasa en el ambiente j . Solo considera personas sin viajes inválidos.

- 1 Definir la matriz auxiliar $C = (c_{ij})_{i=1\dots n, j=1\dots m}$. c_{ij} guardará el tiempo total que las personas de la clase i pasan en el ambiente j . Se inicializa en 0;
 - 2 Definir $T = T_f - T_0$ el tiempo total considerado.
 - 3 Definir $N_i \leftarrow 0$ para cada $i \in 1 \dots n$ para contar el total de personas en cada clase.
 - 4 **para todo** $p \in P$ **hacer**
 - 5 Guardar en i el valor $\mathcal{C}(p)$, la clase de p ;
 - 6 **si** *La persona p no tiene viajes, $V_p = \emptyset$* **entonces**
 - 7 Se supone que p pasa todo su tiempo en el ambiente j_0 por defecto.
 $c_{i, j_0} \leftarrow c_{i, j_0} + T$;
 - 8 Contabilizar una persona válida. $N_i \leftarrow N_i + 1$;
 - 9 **si no, si** V_p *tiene información faltante (propósito, modo, t_0 or t_f de algún $v \in V_p$ no está definido)* **entonces**
 - 10 p no es una persona válida y no es considerada en la matriz final;
 - 11 **en otro caso**
 - 12 $j_{\text{anterior}} \leftarrow j_0$. Ambiente donde terminó el viaje anterior. Comienza en j_0 ;
 - 13 $t_{f_anterior} \leftarrow T_0$. Hora de término del viaje anterior. Vale T_0 por defecto;
 - 14 **para todo** *viaje* $v \in V_p$ **hacer**
 - 15 Obtener propósito, modo, t_0 y t_f asociados al viaje v . Definir
 $j_v \leftarrow \mathcal{J}_m(\text{modo})$ el ambiente asociado al modo del viaje v ;
 - 16 Agregar el tiempo entre el viaje anterior y el viaje v al ambiente donde
 terminó el viaje anterior. $c_{i, j_{\text{anterior}}} \leftarrow c_{i, j_{\text{anterior}}} + t_s - t_{f_anterior}$;
 - 17 Agregar el tiempo del viaje v a su ambiente asociado. $c_{i, j_v} \leftarrow c_{i, j_v} + t_f - t_s$;
 - 18 Actualizar $j_{\text{anterior}} \leftarrow \mathcal{J}_p(\text{propósito})$. El próximo viaje comienza
 donde v terminó.
 - 19 Actualizar $t_{f_anterior} \leftarrow t_f$.
 - 20 **fin**
 - 21 Agregar el tiempo entre el último viaje y el tiempo final T_f v al ambiente
 donde terminó el último viaje. $c_{i, j_{\text{anterior}}} \leftarrow c_{i, j_{\text{anterior}}} + T_f - t_{f_anterior}$;
 - 22 Contabilizar a p como una persona válida. $N_i \leftarrow N_i + 1$;
 - 23 **fin**
 - 24 **fin**
 - 25 Calcular R a partir de C normalizando cada fila. $r_{ij} \leftarrow c_{ij}/TN_i$;
-

Uso del tiempo



Diario de viaje

Figura 2.5: Transformar propósitos de viaje en tiempo de actividad. Fuente: [24], traducción propia.

El algoritmo 1 explica el proceso en detalle de manera general. Cabe mencionar que este algoritmo fue desarrollado de manera independiente y con importantes variantes antes de conocer [24]. Este algoritmo es aplicado para construir una matriz detallada de tiempos de residencia, intentando aprovechar la mayor información posible de los propósitos de viajes. También es usado para construir la matriz de tiempos de residencia con únicamente dos ambientes, pero en ese caso las funciones \mathcal{I}_p , \mathcal{I}_m son considerablemente más sencillas; solo el propósito asociado con ir al hogar se asocia al ambiente **hogar** y todos los demás se asocian a **exterior**. La matriz obtenida usando este algoritmo se denota P^0 . La matriz variable en el tiempo será construida a partir de esta en la siguiente subsección.

2.3.3. Variaciones de la matriz a lo largo del tiempo

Se usan los datos de movilidad obtenidos del Uso de infraestructura de telecomunicaciones, descritos en 2.1.5. Los datos se obtienen del repositorio de Datos-COVID19, descrito en 2.1.6, del Producto 82 específicamente.

Los datos se encuentran en dos archivos en formato `.csv`, correspondientes a la movilidad durante días laborables de la semana y a la movilidad de los fines de semana. Se usa solo el primero de estos. Los atributos incluyen la comuna de la Región Metropolitana con nombre y código único territorial, comuna a la que corresponde la medición, semana epidemiológica de la medición con fecha de inicio y fin, y estimaciones de la variación de movilidad, con cotas inferiores y superiores.

Se utiliza la información de estimaciones de la variación de movilidad (despreciando las cotas inferiores y superiores). Si llamamos m_i^0 a la movilidad de referencia inicial para la comuna i -ésima (la cual es desconocida), y m_i^k a la movilidad de la comuna i -ésima en la semana k -ésima posterior a esa semana de referencia, entonces son reportados los valores p_i^k que satisfacen la ecuación 2.3.

$$m_i^k = p_i^k m_i^0. \quad (2.3)$$

Un detalle a considerar es que es necesario calcular la variación de movilidad para los distintos grupos socioeconómicos, puesto que solo se tiene para las comunas. Si la población de la comuna i -ésima es N_i , e I es un conjunto de comunas, se estima la movilidad grupal con las ecuaciones 2.4. Se busca un parámetro p_I^k tal que 2.5. Bajo el supuesto $m_i^0 := m^0$ para cada i , entonces es fácil ver que se cumple 2.6.

$$m_I^k = \sum_{i \in I} m_i^k N_i, \quad (2.4)$$

$$m_I^k = p_I^k m_I^0, \quad (2.5)$$

$$p_I^k = \sum_{i \in I} p_i^k N_i. \quad (2.6)$$

Se supondrá que la movilidad es constante a lo largo de la semana, incluyendo fines de semana, aunque no se considerarán estos datos. Se define una función $\text{semana}(t)$ que a un tiempo t le asocia la correspondiente semana. Suponemos además que el tiempo que cada clase pasa en el ambiente **exterior** varió de manera proporcional a la movilidad. Eso permite calcular la columna **exterior** de la matriz de tiempos de residencia para cada grupo socioeconómico I con las ecuaciones 2.7. Finalmente, el tiempo que no se ocupa en el ambiente **exterior** se pasa en el ambiente **hogar**, por lo que se cumple 2.8.

$$r_{I,\text{exterior}}(t) = p_I^{\text{semana}(t)} r_{I,\text{exterior}}^0, \quad (2.7)$$

$$r_{I,\text{hogar}}(t) = 1 - r_{I,\text{exterior}}(t). \quad (2.8)$$

2.4. Estimación de parámetros

En la subsección 2.4.1 se elige el filtro de kalman a utilizar. A continuación 2.4.2 detalla las técnicas específicas de filtro de kalman usadas en la estimación del factor sanitario. Luego 2.4.3 trata con la elección del valor β_{exterior} . Posteriormente 2.4.4 detalla cómo se hace la elección de las varianzas del proceso, para obtener estimaciones con sentido. La subsección 2.4.5 habla muy brevemente de la observabilidad del sistema. Finalmente, 2.4.6 da los detalles específicos del código escrito para resolver el problema.

2.4.1. Elección de Filtro de Kalman

Filtro de Kalman [27] es una técnica ampliamente utilizada en ingeniería [28]. Se tiene un sistema cuya evolución está dada por ciertas ecuaciones. El estado del sistema es desconocido pero se cuenta con observaciones ruidosas de este. El objetivo es estimar el estado a partir

de estas observaciones; por ejemplo, estimar la posición de un vehículo cuando solo se conoce su velocidad.

Para tratar con sistemas continuos, como los que resultan al tratar con ecuaciones diferenciales, hay dos alternativas: por una parte, es posible discretizar el sistema mediante métodos como Euler o Runge-Kutta, para luego tratar la no linealidad (este método es llamado discreto-discreto en [47]). Por otra parte, es posible linealizar el sistema primero, lo que da lugar al filtro continuo-discreto.

Tratar con la no linealidad del sistema puede hacerse de varias formas. La más directa es simplemente linealizar las funciones de la dinámica, lo que da lugar al Filtro de Kalman Extendido (EKF por sus siglas en inglés). Hay enfoques más sofisticados que funcionan mejor para problemas altamente no lineales como el *Unscented Kalman Filter* (UKF, Filtro de Kalman “sin olor”) o el *Cubature Kalman Filter* (CDF) que se basan estimar la media y varianza de una distribución a la que es aplicada una función no lineal, mediante la evaluación de esa función en varios puntos estratégicamente elegidos, llamados σ -puntos [66][49].

Las distintas variantes del Filtro de Kalman se han usado previamente en epidemiología para estimación de parámetros, tanto fijos como cambiantes en el tiempo, trabajando con datos relacionados a SIDA [67], ébola [68], dengue [69], tuberculosis [70], entre otros.

Con respecto al uso en COVID-19, se han usado distintas variantes del filtro con distintos propósitos. Los artículos [71] y [72] usan EnKF para hacer predicción. Se han propuesto varios métodos que usan distintas variantes del filtro para estimar el número reproductivo efectivo como [29][73][74]. En [75] se usa EKF para desarrollar un control robusto basado en distanciamiento social, hospitalización y tratamiento, y vacunación. En [76] se usa *Fractional Order Calculus*, un técnica de filtraje de la que EKF puede verse como un caso particular, para estimar transmisibilidad. La referencia [30] propone un método basado en EKF para estimar parámetros variables en el tiempo, usando descenso de gradiente con una función de costo dada por una medida de verosimilitud. Usa el método para estimar un parámetro de transmisibilidad y el número reproductivo efectivo.

Algunos artículos que utilizan técnicas más similares a la metodología propuesta son [77] y [25], que estiman un estado aumentado por la tasa de transmisión (y algunos parámetros extra) mediante EKF discreto-discreto. Por simplicidad de la implementación y a la hora de chequear observabilidad local, también se decidió utilizar EKF en su versión discreta-discreta. Muchos sin embargo usan discretización de Euler; se decidió usar un esquema de Runge-Kutta por estabilidad numérica.

Debido a que se trabajará con datos pasados y no se busca hacer estimación en línea ni predicción, se utiliza suavizado en lugar de filtraje, ver la sección 1.2.4. El suavizado será de intervalo fijo y se usará la versión más conocida, el *Rauch, Tung y Striebel Smoother* o *RTS Smoother* [50][49]. Esto permite obtener mejores estimaciones, pero solo es posible debido a que disponemos de todas las observaciones para un intervalo de tiempo fijo, y que deseamos la mejor estimación posible dadas esas observaciones. Para estimación en línea solo se cuenta con las mediciones hasta el intervalo donde se quiere estimar, no hay observaciones futuras, por lo que se espera tener menor calidad.

2.4.2. Estimación de la tasa de contagios

Una vez elegido la variante de filtro de Kalman a utilizar, es necesario aplicarlo al modelo elegido. Esto se hará en dos etapas: en primer lugar, se trabaja con un modelo sencillo con una única clase y sin estructura de ambientes. En la segunda etapa se trabaja con el modelo multiclase multiambiente.

Ambas etapas siguen un orden similar. En primer lugar se plantea el modelo como una ecuación diferencial con ruido independiente. Después se agrega una variable de casos acumulados y se aumenta el estado con el factor sanitario α . Se discretiza la dinámica usando un esquema de Runge-Kutta de orden 4, obteniendo un sistema discreto no lineal. Después se calcula el jacobiano de ese sistema, el cual debe considerar no solo el modelo SEIR, sino que también el esquema de discretización. A continuación se obtiene una matriz de observación de los casos acumulados, y se amplía esa matriz para imponer la restricción de población constante mediante la técnica de observaciones casi perfectas.

Finalmente se aplica EKF, usando como datos de observaciones el Producto 15 de [35] “Casos nuevos por fecha de inicio de síntomas por comuna”, adecuadamente tratado, es decir, calculando casos acumulados en lugar de casos diarios, y utilizando una media móvil para suavizar los datos. Se puede aplicar además *RTS smoother*.

Siguiendo la misma línea que [29][77][78], supondremos que γ_E y γ_I son constantes y que $\alpha(t)$ es variable y depende de las distintas medidas sanitarias implementadas.

Modelo con una clase

Para fijar ideas, se trabaja con el modelo SEIR clásico de una sola clase, incorporando ruido, como en el sistema de ecuaciones 2.9. w_1, \dots, w_4 son procesos brownianos independientes entre sí, ponderados por constantes g_1, \dots, g_4 .

$$\begin{aligned} S'(t) &= -\alpha(t)S(t)I(t) + g_1w_1(t) \\ E'(t) &= \alpha(t)S(t)I(t) - \gamma_E E(t) + g_2w_2(t) \\ I'(t) &= \gamma_E E(t) - \gamma_I I(t) + g_3w_3(t) \\ R'(t) &= \gamma_I I(t) + g_4w_4(t). \end{aligned} \tag{2.9}$$

$$C(t) = \int_{t_0}^t \gamma_E E(s) ds. \tag{2.10}$$

Se supondrá que γ_E y γ_I son conocidos, y se busca estimar $\alpha(t)$ a partir de observaciones ruidosas de los casos infectados acumulados, los cuales se calculan usando la expresión 2.10. Se trabajará con un modelo aumentado, agregando las ecuaciones 2.11 a la dinámica. 2.11a permite guardar los casos acumulados en un estado, lo que facilita plantear la ecuación de observación. 2.11b agrega el factor sanitario α a la dinámica sin ser modificada por ella, de forma que solo es actualizada por el paso de análisis del filtro de Kalman.

$$C'(t) = \gamma_E E(t) \quad (2.11a)$$

$$\alpha'(t) = 0. \quad (2.11b)$$

Llamando $\mathbf{x} = (S, E, I, R, C, \alpha)^\top$, se obtiene la ecuación de estado 2.12, con parámetros conocidos $p = (\gamma_E, \gamma_I)$. \mathbf{G} es una matriz diagonal que pondera el ruido, y \mathbf{w} un vector de brownianos independientes. Si bien f_p no depende del tiempo de manera explícita, al trabajar con el modelo completo si habrá dependencia, por lo que se decidió escribir $f_p(\mathbf{x}, t)$ en lugar de simplemente $f_p(\mathbf{x})$.

$$\mathbf{x}'(t) = \begin{pmatrix} S \\ E \\ I \\ R \\ C \\ \alpha \end{pmatrix}' = \begin{pmatrix} -\alpha(t)S(t)I(t) \\ \alpha(t)S(t)I(t) - \gamma_E E(t) \\ \gamma_E E(t) - \gamma_I I(t) \\ \gamma_I I(t) \\ \gamma_E E(t) \\ 0 \end{pmatrix} + \mathbf{G}w(t) = f_p(\mathbf{x}, t) + \mathbf{G}w(t). \quad (2.12)$$

Para utilizar el filtro discreto-discreto, se requiere discretizar la dinámica f_p , para lo que se usa un esquema de Runge Kutta de orden 4. Para un instante de tiempo t_k , un intervalo de tiempo Δt y un estado \mathbf{x}_k , y definiendo $\Delta \mathbf{x}_1, \dots, \Delta \mathbf{x}_4$ como en 2.13, 2.14 permite obtener una estimación 2.15 para \mathbf{x}_{k+1} . Esto corresponde a la ecuación de estado discretizada, la cual es no lineal. Notamos que $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$ y $\mathbf{Q}_k = \frac{\mathbf{Id}}{\sqrt{\Delta t}}$. Una explicación de esto puede encontrarse en el capítulo 8 de [49].

$$\begin{aligned} t_{n+1/2} &:= t_n + \Delta t/2 \\ \Delta x_1 &= f(x(t_n), t_n) \\ \Delta x_2 &= \Delta t f(x(t_n) + \Delta x_1, t_{n+1/2}) \\ \Delta x_3 &= \Delta t f(x(t_n) + \Delta x_2, t_{n+1/2}) \\ \Delta x_4 &= \Delta t f(x(t_n) + \Delta x_3, t_{n+1}) \end{aligned} \quad (2.13)$$

$$\mathbf{x}_k + \frac{1}{6} \left(\Delta \mathbf{x}_1 + 2\Delta \mathbf{x}_2 + 2\Delta \mathbf{x}_3 + \Delta \mathbf{x}_4 \right) =: \mathcal{F}_{p, \Delta t}(\mathbf{x}_k, t_k). \quad (2.14)$$

$$\mathbf{x}_{k+1} = \mathcal{F}_{p, \Delta t}(\mathbf{x}_k, t_k) + \mathbf{G}w_k. \quad (2.15)$$

Ahora es necesario trabajar con la no linealidad, lo que se hará mediante el Filtro de Kalman Extendido, expuesto en la sección 1.2.2. Para eso, es necesario calcular el jacobiano de la función $\mathcal{F}_{p, \Delta t}$. Múltiples usos de la regla de la cadena más la definición del esquema de RK4 permiten obtener las ecuaciones 2.16.

$$\begin{aligned}
D_x(\mathcal{F}_{p,\Delta t})(x_k, t_k) &= I + \frac{1}{6} \left(D_x h_1 + 2D_x h_2 + 2D_x h_3 + D_x h_4 \right) \\
t_{k+1/2} &:= t_k + \Delta t/2 \\
h_1 &= f(x_k, t_k) \\
D_x h_1 &= D_x f(x_k, t_k) \\
h_2 &= \Delta t f(x_k + h_1, t_{k+1/2}) \\
D_x h_2 &= \Delta t D_x f(x_k + h_1, t_k) (I + D_x h_1) \\
h_3 &= \Delta t f(x_k + h_2, t_{k+1/2}) \\
D_x h_3 &= \Delta t D_x f(x_k + h_2, t) (I + D_x h_2) \\
h_4 &= \Delta t f(x_k + h_3, t_{k+1}) \\
D_x h_4 &= \Delta t D_x f(x_k + h_3, t) (I + D_x h_3).
\end{aligned} \tag{2.16}$$

La ecuación de observación 2.17 consiste, en primera instancia, simplemente en una matriz que obtiene la coordenada de los casos acumulados C dentro del vector \mathbf{x} . La observación se obtiene con ruido, dado por $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$.

$$\mathbf{y}_k = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{x}_k + \mathbf{v}_k. \tag{2.17}$$

Debe mencionarse que el sistema no lineal tiene restricciones; por una parte, las variables de estado deben ser positivas para que tengan sentido. Esto se consigue aplicando la función $x \mapsto \max(x, 0)$ (por coordenada). Por otra parte, al elegir $b_i = d_i = 0$ se impone que el total de la población debe mantenerse constante. Para conseguir esto se utiliza la técnica de observaciones casi perfectas, explicada en la subsección 1.2.5. Se elige usar una restricción suave para imponer $S(t) + E(t) + I(t) + R(t) \approx N$, con lo que se obtiene la ecuación de observación 2.18 definitiva.

$$\begin{bmatrix} \mathbf{y}_k \\ N \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \end{bmatrix} \mathbf{x}_k + \begin{bmatrix} \mathbf{v}_k \\ \varepsilon \end{bmatrix} = \begin{bmatrix} C_k \\ S_k + E_k + I_k + R_k \end{bmatrix} + \begin{bmatrix} \mathbf{v}_k \\ \varepsilon \end{bmatrix}. \tag{2.18}$$

Finalmente, a modo de resumen; se utiliza el Filtro de Kalman Extendido sobre el sistema dado por 2.15. Este sistema fue obtenido extendiendo un modelo SEIR con la variable C de casos acumulados, y con un factor α de la tasa de contagio, y discretizado mediante el esquema RK4. El sistema es no lineal en el estado, y lineal en el ruido. La linearización se obtiene a partir del jacobiano calculado en 2.16. Para asegurar la positividad de las variables se aplica la función $\max(x, 0)$ después de cada paso de *forecast* y análisis. Para imponer que la población se mantenga constante, se usa la técnica de observaciones casi perfectas de [51]. La ecuación de observación final es 2.18, la cual es lineal.

Una vez calculadas las iteraciones de filtro de Kalman, es posible obtener los resultados suavizados con el *RTS smoother* mencionado en la sección 1.2.4. Los detalles de la implementación se encuentran en la subsección 2.4.6.

Modelo multiclase multiambiente

Al trabajar con el modelo con una clase ya se ha tratado la mayor parte del problema. Ahora se extiende esa metodología al modelo con n clases y m ambientes presentado en la sección 2.2. El vector de variables de estado es $\mathbf{x} = (S_1, \dots, S_n, E_1, \dots, E_n, I_1, \dots, I_n, R_1, \dots, R_n)$. Se supone que β_1, \dots, β_m son conocidos y fijos. Se supondrá además que $\gamma_{Ei} := \gamma_E, \gamma_{Ii} := \gamma_I$, es decir, que las tasas de incubación y recuperación son las mismas para todas las clases. Se estudiarán los efectos de esto más adelante, puesto que es una simplificación importante. Las ecuaciones para $i \in 1 \dots n$ están en 2.19.

$$S'_i(t) = -\lambda_i(\mathbf{x}, t)S_i(t) + g_i^S w_i^S(t) \quad (2.19a)$$

$$E'_i(t) = \lambda_i(\mathbf{x}, t)S_i(t) - \gamma_E E_i(t) + g_i^E w_i^S(t)$$

$$I'_i(t) = \gamma_E E_i(t) - \gamma_I I_i(t) + g_i^I w_i^S(t)$$

$$R'_i(t) = \gamma_I I_i(t) + g_i^R w_i^S(t)$$

$$\lambda_i(\mathbf{x}, t) = \alpha_i(t) \sum_{j=1}^m \beta_j p_{ij}(t) \left(\frac{\sum_{k=1}^n p_{kj}(t) I_k}{\sum_{k=1}^n p_{kj}(t) N_k} \right). \quad (2.19b)$$

Al igual que en el caso con una clase, se agregan las ecuaciones 2.21, correspondientes a las dinámicas de los casos acumulados 2.21a, calculados según 2.20, y de los factores sanitarios $\alpha_i(t)$ 2.21b.

$$C_i(t) = \int_{t_0}^t \gamma_E E_i(s) ds, \quad (2.20)$$

$$C'_i(t) = \gamma_E E_i(t) \quad (2.21a)$$

$$\alpha'_i(t) = 0. \quad (2.21b)$$

Como antes, discretizando con un esquema de Runge Kutta podemos obtener un sistema de la forma 2.22. Para tratar con la no linealidad basta con usar las mismas ecuaciones 2.16 calculadas previamente para el jacobiano.

$$\mathbf{x}_{k+1} = \mathcal{F}_{p, \Delta t}(\mathbf{x}_k, t_k) + \mathbf{G} \mathbf{w}_k. \quad (2.22)$$

La ecuación de observaciones en este caso es similar a 2.17. Definiendo $\mathbf{0}_n$ y \mathbf{Id}_n como la matriz de 0's y la matriz identidad de tamaño $n \times n$, entonces la ecuación de observaciones está dada por 2.23.

$$\mathbf{y}_k = [\mathbf{0}_n \quad \mathbf{0}_n \quad \mathbf{0}_n \quad \mathbf{0}_n \quad \mathbf{Id}_n \quad \mathbf{0}_n] \mathbf{x}_k + \mathbf{v}_k. \quad (2.23)$$

Denotando $\mathbf{N} := (N_1, \dots, N_n)^\top$ y $\boldsymbol{\varepsilon} := (\varepsilon, \dots, \varepsilon)^\top$, las ecuaciones de observación que imponen además que $S_i(t) + E_i(t) + I_i(t) + R_i(t) \approx N_i$ están en 2.24. $S_{i,k}$ denota la aproximación de susceptibles del sistema discreto, para la clase i -ésima, a tiempo t_k , y lo mismo para $C_{i,k}$.

$$\begin{bmatrix} \mathbf{y}_k \\ \mathbf{N} \end{bmatrix} = \begin{bmatrix} \mathbf{0}_n & \mathbf{0}_n & \mathbf{0}_n & \mathbf{0}_n & \mathbf{Id}_n & \mathbf{0}_n \\ \mathbf{Id}_n & \mathbf{Id}_n & \mathbf{Id}_n & \mathbf{Id}_n & \mathbf{0}_n & \mathbf{0}_n \end{bmatrix} \mathbf{x}_k + \begin{bmatrix} \mathbf{v}_k \\ \boldsymbol{\varepsilon} \end{bmatrix} = \begin{bmatrix} C_{1,k} \\ \vdots \\ C_{n,k} \\ S_{1,k} + E_{1,k} + I_{1,k} + R_{1,k} \\ \vdots \\ S_{n,k} + E_{n,k} + I_{n,k} + R_{n,k} \end{bmatrix} + \begin{bmatrix} \mathbf{v}_k \\ \boldsymbol{\varepsilon} \end{bmatrix}. \quad (2.24)$$

2.4.3. Estimación de parámetros fijos

Hasta ahora se ha supuesto que γ_E, γ_I y β_1, \dots, β_m son valores fijos y conocidos. Los valores γ_E y γ_I pueden ser estimados a partir de la literatura, pero no es tan claro cómo estimar los riesgos específicos por ambiente. Puesto que se trabajará solamente con $m = 2$ (**hogar** y **exterior**), se fijará el valor de $\beta_{\text{hogar}} = 1$, por lo que basta con dar un valor a β_{exterior} . Lo esperado es que estar en el exterior sea más riesgoso que en el hogar, así que tiene sentido considerar $\beta_{\text{exterior}} > \beta_{\text{hogar}}$.

Para darle un valor a este parámetro, se considerará la expresión 2.25, que corresponde a la contribución j -ésima a λ_i , dado por la ecuación 2.19b. Se buscarán valores en concordancia con [65][19], quienes afirman que 1/3 de los contagios ocurren en el hogar. La sensibilidad a este parámetro se estudia en 3.2.2.

$$p_{ij} \frac{\sum_{k=1}^n I_k p_{kj}}{\sum_{k=1}^n N_k p_{kj}}. \quad (2.25)$$

2.4.4. Matrices de covarianza: ajuste por caso sintético

Ni [29] ni [25], quienes usan métodos similares, explican cómo elegir las matrices de ruido, y las dejan como parámetros a ajustar. Al hacer pruebas con la metodología propuesta, se verificó que los resultados pueden depender considerablemente de esta decisión, por lo que se decidió hacer un ajuste por medio de un caso sintético.

Se eligen valores fijos γ_I, γ_E y β_{exterior} . Se resolverán las ecuaciones diferenciales del sistema epidemiológico SEIR dado por 2.2, utilizando la matriz de tiempos de residencia calculada en la sección 2.3. Para eso se fabrican funciones $\alpha_i(t)$ sintéticas, de manera que se obtenga alguna solución interesante (un doble *peak* en los infectados por ejemplo) y que los órdenes de magnitud de los casos acumulados obtenidos $C_i(t)$ sean similares a los que se usarán con datos reales.

Posteriormente, se hacen varias pruebas de la metodología propuesta, suponiendo condición inicial y parámetros desconocidos, para intentar estimar los estados. Las matrices de covarianza iniciales y los ponderadores del ruido en el modelo se ajustan de tal forma que las estimaciones obtenidas de los estados se encuentren a una desviación estándar de los estados de la solución real, y se eligen como una fracción de las condiciones iniciales $S_{i,0}$, de forma que sean aplicables en una mayor variedad de casos. Estos valores se usarán en el caso real.

2.4.5. Observabilidad

Como ya vimos, en el caso lineal con dinámica constante, se requiere observabilidad para tener convergencia al estado real del sistema, pero el caso no lineal es más difícil. En este trabajo solamente se verifica la observabilidad localmente para el sistema linealizado, de manera numérica.

2.4.6. Implementación

Algunos artículos que siguen metodologías similares y cuyos códigos están disponibles son [29] y [25]. Ambos están escritos en Matlab y no utilizan ninguna librería de Filtro de Kalman, sino que este es programado desde cero para cada caso.

A pesar de que Matlab es un lenguaje dominante en computación científica, se decidió utilizar Julia. Entre la serie de ventajas que proporciona utilizar este lenguaje se encuentran; en primer lugar es compilado, al igual que lenguajes como C, C++ y Java, por lo que si el código está adecuadamente escrito es considerablemente más rápido que lenguajes interpretados como Matlab o Python. En general, cuando es posible vectorizar el código esta diferencia no es significativa, pero eso no es posible completamente en un método iterativo como Filtro de Kalman.

En segundo lugar, la sintaxis en Julia es similar a la de lenguajes de alto nivel como Matlab y Python, por lo que es simple de escribir, a diferencia de lenguajes más verbosos donde se deben declarar explícitamente los tipos como C o Java.

En tercer lugar, Julia cuenta con dos librerías que ofrecen características de interés para este trabajo. La primera, `DifferentialEquations.jl` [79], permite acceder a una amplia variedad de *solvers* para distintos tipos de ecuaciones diferenciales, a través de una única interfaz. La segunda, `ModelingToolkit.jl` [80], que aún no ha alcanzado su versión `v1.0`, provee de un sistema simbólico de modelado basado en ecuaciones, que está integrado con `DifferentialEquations.jl` y que optimiza el código para obtener mayor rendimiento. Esto facilita el cálculo de jacobianos y la resolución de ecuaciones diferenciales a partir de un conjunto de ecuaciones simbólicas.

A diferencia de [29] y [25], se decidió escribir código que pudiera ser reutilizado y extendido con mayor facilidad, y que contara con interfaces comunes que permitieran utilizar diferentes tipos de filtros de Kalman con cambios mínimos para el usuario final. Estas directrices guiaron la construcción de `KalmanFilter.jl`, que se encuentra disponible en el

repositorio en GitHub <https://github.com/tabitaCatalan/kalman>. El código específico para la estimación de parámetros se encuentra también disponible en el repositorio <https://github.com/tabitaCatalan/CovidMTK>.

2.5. Evaluación del modelo mediante casos hipotéticos

Para evaluar los resultados obtenidos con la metodología propuesta se usarán los resultados obtenidos en la sección anterior (matriz de tiempos de residencia, factor sanitario) para generar varias situaciones hipotéticas.

En la sección 2.3 se calculó una matriz de tiempos de residencia variable en el tiempo $(p_{i,j})_{i \in 1..5, j \in \{\text{hogar}, \text{exterior}\}}$. Una vez aplicada la metodología de la sección anterior 2.4 se obtiene una estimación de los factores sanitarios $(\alpha_i(t))_{i \in 1..5}$ para cada grupo. En este apartado se busca evaluar, utilizando variantes de esas estimaciones, una serie de casos hipotéticos. Se considerarán tres niveles de cuarentena y tres niveles de cuidado, definidos a continuación. Se recuerda que para m ambientes, basta con definir los valores de la matriz de residencia en $m - 1$ columnas, debido a la restricción $\sum_{j=1}^m p_{ij} = 1$.

- **Cuarentena normal:** todos se comportan de la misma forma. Se utiliza la matriz de tiempos de residencia variable en el tiempo que se usó antes.

$$(p_{\text{normal}})_{i,\text{hogar}}(t) = p_{i,\text{hogar}}(t)$$

- **Cuarentena fuerte:** cada grupo en el hogar tanto tiempo como es posible. Se utiliza en cada t el valor máximo de entre todos los grupos.

$$(p_{\text{cuarentena fuerte}})_{i,\text{hogar}}(t) = \max_{i \in 1..n} \{(p_{\text{normal}})_{i,\text{hogar}}(t)\}$$

- **Sin cuarentena:** durante pandemia no se realiza ningún tipo de cuarentena ni reducción de movilidad. Todos conservan los tiempos de residencia en condiciones normales.

$$(p_{\text{sin cuarentena}})_{i,\text{hogar}}(t) = (p_{\text{normal}})_{i,\text{hogar}}(0)$$

- **Cuidado normal:** todos se comportan de la misma forma. Se utilizan los factores sanitarios estimados en la sección anterior

$$(\alpha_{\text{normal}})_i(t) = \alpha_i(t)$$

- **Cuidado insuficiente:** todos tiene el factor sanitario de la clase más vulnerable, $i = 5$.

$$(\alpha_{\text{descuido}})_i(t) = \alpha_5(t)$$

- **Cuidado extra:** todos se comportan como la clase que en general tuvo el factor sanitario más bajo. En este caso, esa clase corresponde a $i = 1$, la clase más acomodada.

$$(\alpha_{\text{cuidado extra}})_i(t) = \alpha_1(t)$$

Capítulo 3

Resultados

En este capítulo se exponen los resultados obtenidos con la metodología propuesta. La sección 3.1 está dedicada a la matriz de tiempos de residencia. Esta se estimó en dos etapas; en primer lugar, usando la Encuesta Origen-Destino tratada mediante el algoritmo 1 para obtener el comportamiento de los santiaguinos en condiciones normales. A continuación, se utilizan datos de movilidad para actualizar esta matriz, de forma que refleje las variaciones en pandemia.

En la sección 3.2 se muestra la estimación del factor sanitario. Esto se hace en dos etapas; en primer lugar se calcula para un caso sintético, donde la solución buscada está disponible. El conjunto de parámetros que permite una buena recuperación del factor sanitario se extrapola al caso real, donde se utilizan los casos confirmados obtenidos de Datos-COVID19 presentados en la subsección 2.1.6. Se estudia además la sensibilidad de las estimaciones ante distintos parámetros.

Finalmente, en la sección 3.3, se usan los valores estimados en la sección anterior para simular varios casos hipotéticos; ¿cómo habría sido el desarrollo de la pandemia si todos se cuidaran de la misma forma pero además pudieran guardar una cuarentena más estricta? ¿y si no se hacía cuarentena pero todos respetaban las medidas sanitarias de una mejor manera?. Se estudian dos escenarios dados por el parámetro β_{exterior} .

3.1. Matriz de tiempos de residencia

En esta sección se muestran las matrices de tiempos de residencia estimadas para la ciudad de Santiago. La estimación se hizo en dos etapas; en primer lugar se calcula una matriz en condiciones normales, antes de la llegada del virus, lo que se presenta en la subsección 3.1.1. Preliminarmente se calculó una matriz muy granular, la cual es presentada pues es de interés en si misma. La versión a utilizar, sin embargo, es mucho más simple.

En la segunda etapa de la estimación, se utilizan datos de movilidad para modificar la matriz de condiciones normales, a fin de que refleje el comportamiento de la población en pandemia. Esto se presenta en la subsección 3.1.2.

3.1.1. Santiago en condiciones normales

Originalmente se deseaba trabajar con una mayor cantidad de clases y ambientes. La idea era utilizar una clasificación por sexo, edad e ingresos familiares, e incluir una lista bastante extensa de ambientes; prácticamente todos los propósitos incluidos en la Encuesta Origen-Destino. Es por esto que se calculó una matriz granular de tiempos de residencia en condiciones normales, mediante el algoritmo 1. Los resultados son presentados en 3.1.1.

Esta matriz, sin embargo, es demasiado detallada. Como se discutió en 2.3.1, no se cuenta con los datos necesarios para actualizarla sin hacer una gran cantidad de supuestos, por lo que solamente se utilizarán $n = 5$ clases y $m = 2$ ambientes, hogar y exterior. Los resultados se muestran en 3.1.1.

El código en Matlab para esta sección se encuentra disponible en el repositorio de GitHub <https://github.com/tabitaCatalan/lagrangian-time>, más específicamente en el directorio `src/time_residence_matrix/`.

Matriz detallada

Se aplica el algoritmo 1 a la Encuesta Origen-Destino Santiago 2012, ya descrita en 2.1.3. Las clases son elegidas combinando los tres criterios de la tabla 3.1; tres niveles de edad, sexo y tres niveles socioeconómicos dan lugar a 18 clases. El nivel socioeconómico es calculado en base a los datos de ingreso de cada persona provistos por la encuesta, agrupando a nivel de hogar y dividiendo por la cantidad de habitantes del hogar. Los ingresos de corte son elegidos de tal forma que cada tramo socioeconómico tiene unas 20 000 personas.

Edad (años)		Sexo	Nivel Socioeconómico (ingreso per cápita)	
Joven	0-24	Hombre	Bajo	\leq \$111 666
Adulto	25-64	Mujer	Medio	\$ 111 667 - \$ 199 999
Mayor	65 o más		Alto	\geq \$ 200 000

Tabla 3.1: Criterios usados para obtener las clases de la matriz detallada a partir de la EOD2012 Santiago.

Los ambientes utilizados están basados en la lista de propósitos de viajes y los modos de transporte usados. Son los siguientes: hogar, trabajo, estudios, compras, visitas, salud, trámites, recreación, transporte público, auto, caminata, bicicleta y otros. La lista de propósitos de viajes y sus ambientes asociados se encuentra en la tabla 3.2. La lista de modos está en la tabla 3.3.

La matriz obtenida se presenta en la figura 3.1. Como era de esperarse, todas las clases pasan la mayor parte de su tiempo en el hogar, de hecho, se redujo el tiempo en el hogar en 7 horas, asociadas al de tiempo de sueño, y se volvió a normalizar la matriz por filas, de forma que las proporciones de tiempo por clase siguieran sumando 1.

De esta matriz pueden hacerse algunas observaciones. Los adultos pasan una cantidad

Propósito de viaje	
1. Al trabajo	trabajo
2. Por trabajo	trabajo
3. Al estudio	estudios
4. Por estudio	estudios
5. De salud	salud
6. Visitar a alguien	visitas
7. Volver a casa	hogar
8. Buscar o dejar a alguien	visitas
9. Comer o tomar algo	compras
10. Buscar o dejar algo	compras
11. De compras	compras
12. Tramites	trámites
13. Recreación	recreación
14. Otra actividad	otros

Tabla 3.2: Propósitos de viaje de la EOD2012 y sus ambientes asociados, lo que permite definir \mathcal{J}_p .

Modo de transporte	
1:Auto	auto
2:Bus TS	transporte público
3:Bus no TS	transporte público
4:Metro	transporte público
5:Taxi Colectivo	transporte público
6:Taxi	auto
7:Bus TS - Bus no TS	transporte público
8:Auto - Metro	transporte público
9:Bus TS - Metro	transporte público
10:Bus no TS - Metro	transporte público
11:Taxi Colectivo - Metro	transporte público
12:Taxi - Metro	transporte público
13:Otros - Metro	transporte público
14:Otros - Bus TS	transporte público
15:Otros - Bus TS - Metro	transporte público
16:Otros	otros
17:Caminata	caminata
18:Bicicleta	bicicleta

Tabla 3.3: Modos de transporte de la EOD2012 y sus ambientes asociados, lo que permite definir \mathcal{J}_m .

considerable de tiempo en el trabajo, en especial los hombres. Los jóvenes pasan mucho tiempo en el ambiente estudios, especialmente los de clase alta; los jóvenes con menos ingresos pasan más tiempo en el trabajo. Los adultos mayores pasan mucho más tiempo en el hogar, pero los hombres pasan una cantidad importante de tiempo en el trabajo.

El nivel socioeconómico medio pasa menos tiempo de compras y de visita que los demás. Los adultos mayores son quienes pasan más tiempo en el ambiente salud, especialmente los de clase baja. El auto es bastante más utilizado por la clase alta y por los adultos hombres de clase media, mientras que el transporte público es más predominante en adultos de la clase media y baja.

En general los resultados son los esperados para una ciudad con una segregación importante como lo es Santiago. Es posible que en la última década se haya reducido la brecha de género, pero eso queda fuera del alcance de este trabajo.

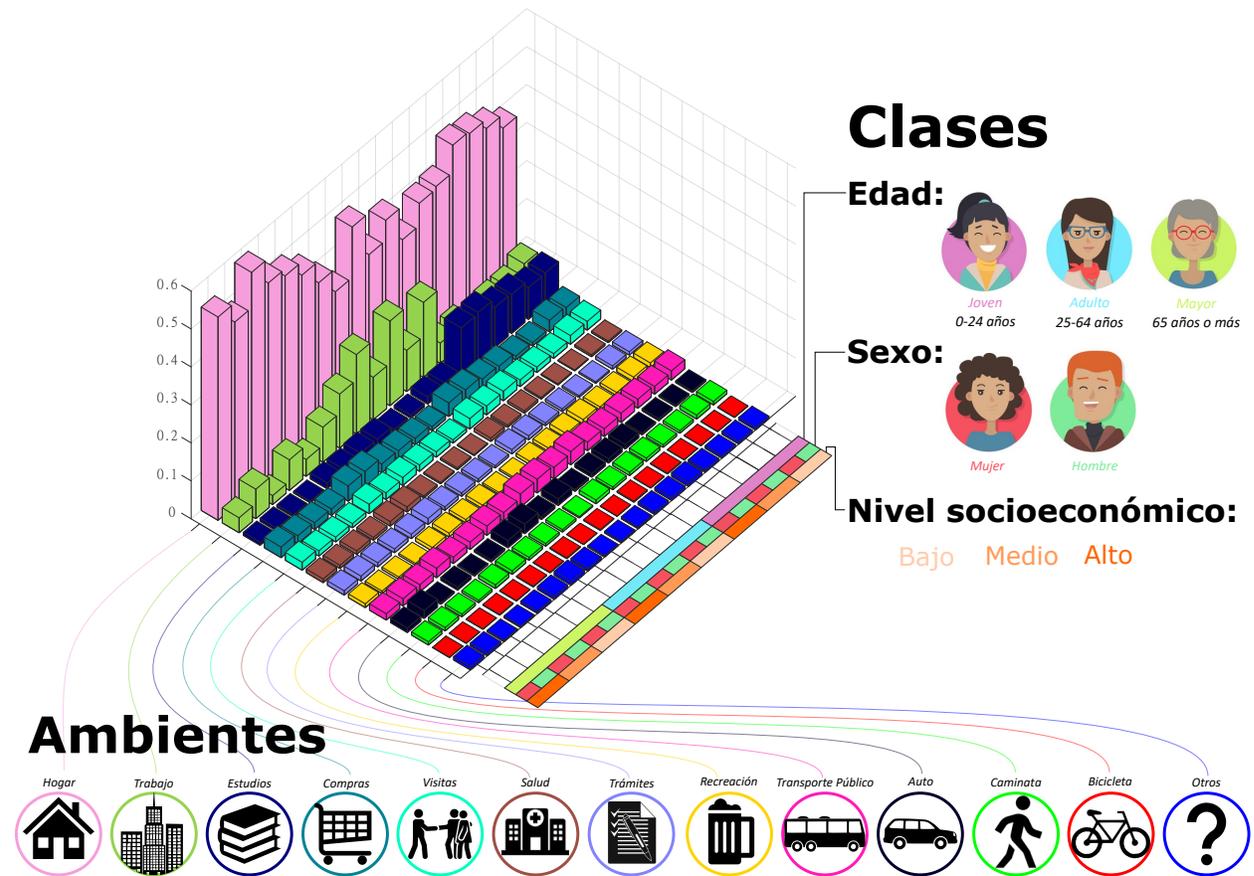


Figura 3.1: Matriz detallada de tiempos de residencia para Santiago. Considera clases basadas en nivel socioeconómico (según ingreso promedio del hogar), edad y sexo. 13 ambientes, basados en los propósitos de viajes de la encuesta. Se restan 7 horas del tiempo en el hogar (tiempo de sueño) y se normalizan las filas para que sumen 1.

Matriz para las simulaciones

La aplicación del algoritmo tomando en consideración las decisiones tomadas en la sección 2.3.1, específicamente la elección de 5 clases agrupadas por IPS como se detalla en la tabla 2.1, y del uso de los ambientes hogar y exterior. Se obtiene una matriz de tiempos de residencia P^0 mucho más sencilla, presentada en la tabla 3.4.

Clase	Prioridad Social	$P^0_{:,hogar}$	$P^0_{:,exterior}$
1	Sin Prioridad	0.73	0.27
2	Baja	0.76	0.24
3	Media Baja	0.77	0.23
4	Media Alta	0.77	0.23
5	Alta	0.77	0.23

Tabla 3.4: Matriz de tiempos de residencia para las simulaciones. La clase sin prioridad corresponde a la más acomodada y la clase con prioridad alta a la más vulnerable.

3.1.2. Santiago a lo largo del tiempo

Una vez que una matriz P^0 ha sido calculada, se requiere modificar esta matriz a fin de obtener $P(t)$ que refleje el comportamiento de la población en pandemia. Estas modificaciones se hacen utilizando los datos de movilidad, siguiendo la metodología presentada en 2.3.3; los datos de movilidad comunal son agregados ponderando por la población de la comuna, y suponiendo que la movilidad inicial era la misma en cada grupo.

La figura 3.2 muestra la variación de movilidad para los distintos grupos socioeconómicos con respecto a la movilidad de la semana base. Se observa que la línea para la clase Sin Prioridad, es decir, la más acomodada, se mantiene por debajo de las otras, lo que significa que logró reducir su movilidad mucho más.

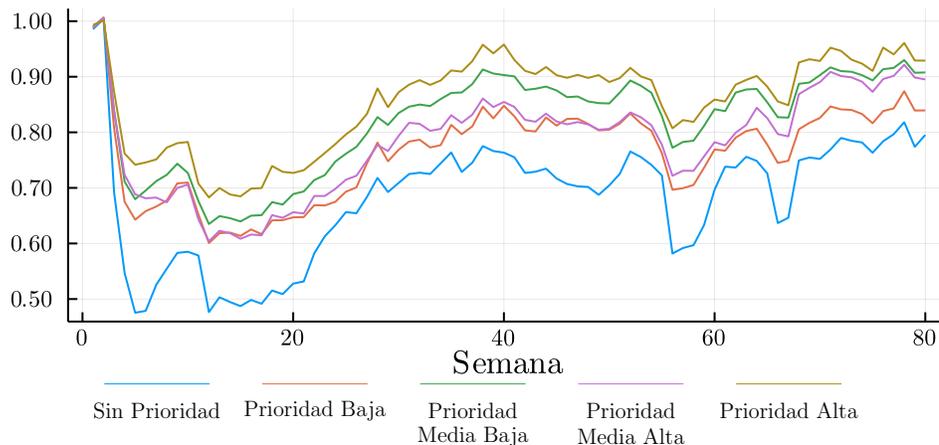


Figura 3.2: Fracción de la movilidad en pandemia por grupos, con respecto a la movilidad normal. Se considera la semana base como la correspondiente a 02/mar/2020 - 06/mar/2020.

Ahora bien, al combinar esos datos con los de 3.1.1 se obtiene la imagen 3.3, que representa el tiempo de residencia estimado en el ambiente **exterior**. Se puede ver que, para la clases más acomodada, el mayor tiempo que pasan en el exterior compensa su mayor reducción de movilidad, haciendo menos notorias las diferencias.

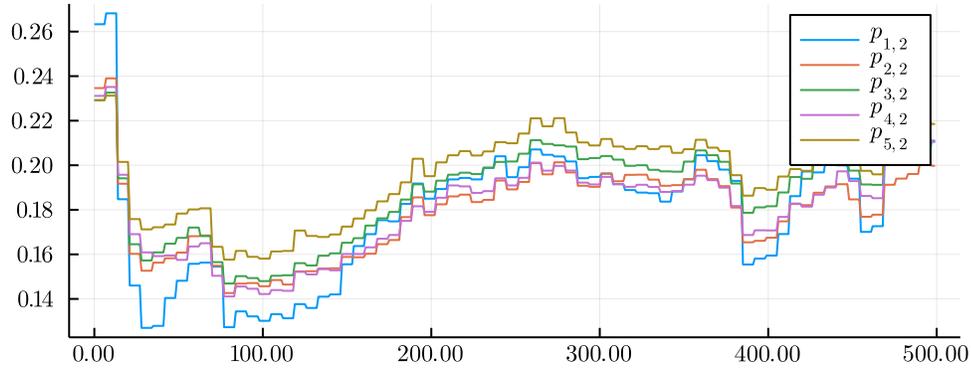


Figura 3.3: Tiempo de residencia en exterior para las 5 clases elegidas, desde 02/mar/2020.

3.2. Estimación de parámetros

Este apartado presenta los resultados de la estimación del factor sanitario. Esto se hace en dos etapas; en primer se calcula para un caso sintético, donde la solución buscada está disponible. Se estudia además la sensibilidad de los resultados ante distintos parámetros. El conjunto de valores que permite una buena recuperación del factor sanitario se extrapola al caso real, donde se utilizan los datos de casos confirmados obtenidos de Datos-COVID19 presentados en la subsección 2.1.6.

Más específicamente, la subsección 3.2.1 muestra la recuperación de valores para un caso sintético. Las subsecciones siguientes se dedican a un estudio de sensibilidad con respecto a los parámetros. En primer lugar, 3.2.2 muestra la influencia del parámetro de riesgo exterior y de las condiciones iniciales del filtro dadas al factor sanitario. En segundo lugar, en la subsección 3.2.3 se exponen los efectos de las matrices de covarianza en los resultados, mostrando varios casos donde los valores no permiten obtener un buen resultado.

Los valores para ajustar el filtro de Kalman en el caso sintético son usados en el modelo que observa datos reales. El factor sanitario estimado para este modelo se encuentran en la subsección 3.2.4. Se hace también un análisis de sensibilidad con respecto a las tasas de incubación y recuperación, el que se encuentra en , 3.2.5. En todos los resultados presentados se utiliza $\Delta t = 1$ [día]. Se probó con valores más pequeños sin alterar el desempeño del método.

3.2.1. Modelo multiclase con datos sintéticos

En este apartado se busca recuperar dos factores sanitarios diferentes usados para generar observaciones sintéticas. El hecho de que la solución es conocida permite estudiar más fácilmente los parámetros del modelo y el filtro. El objetivo es obtener valores que puedan extrapolarse al caso real.

Se utilizan los datos de movilidad de dos comunas (San Bernardo y las Condes) para generar las matrices de tiempos de residencia variables en el tiempo. Se fijan valores $\gamma_E = 1/5.1$, $\gamma_I = 1/8.2$, riesgos $\beta_{hogar} = 1$ y $\beta_{exterior} = 1.88$. Mediante prueba y error se genera

dos controles diferentes para cada clase, de forma que el total de casos acumulados tengan órdenes de magnitud similares a los datos reales.

Estos controles permiten resolver una ecuación diferencial ordinaria y obtener las series de casos acumulados C_1, C_2 . Las observaciones son obtenidas con $\Delta t = 1$ [día]. Se utilizan estas observaciones directamente (sin añadir ruido), y el filtro asume que tienen un ruido pequeño ($G = \text{diag}(10)$).

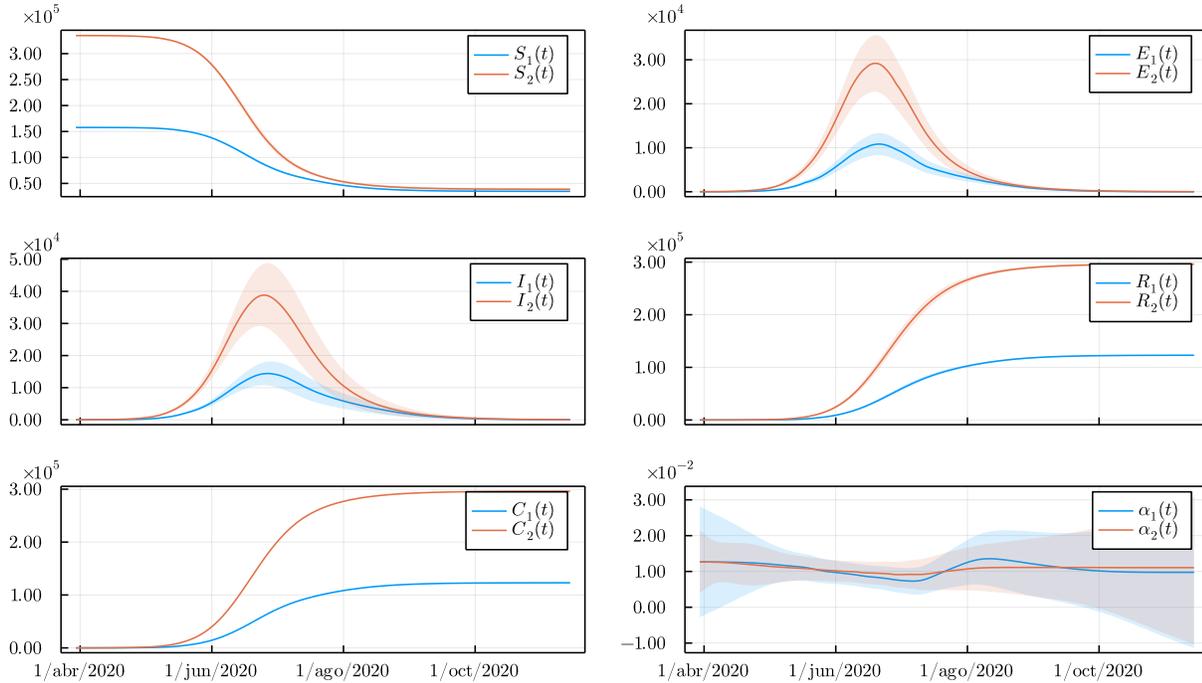
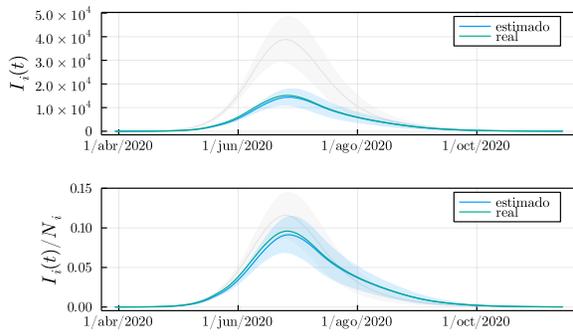


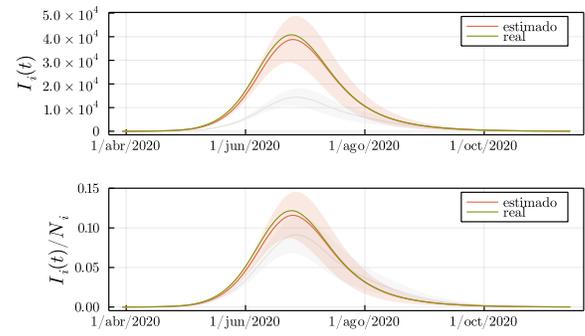
Figura 3.4: Ejemplo de estimación de estados y parámetros variables en el tiempo con *Kalman Smoother*, a partir de las observaciones del número de casos acumulados C_1, C_2 .

Al usar filtro de Kalman se suponen condiciones iniciales y parámetros desconocidos, es decir, no se usan los mismos usados para generar la solución. Los valores de varianza del ruido del proceso son elegidas de modo tal que las soluciones reales quedan a una desviación estándar de la estimación, como se ve en la figura 3.5. Un ejemplo de las estimaciones de estado obtenidas puede verse en la figura 3.4.

Como se ve en la figura 3.6, es posible estimar el factor sanitario a partir de los datos, con ciertas consideraciones. En las secciones siguientes se hablará en más detalle de la sensibilidad de la estimación con respecto a los parámetros del filtro. Más específicamente, se estudia la sensibilidad de la estimación del factor sanitario con respecto al parámetro β_{exterior} y la condición inicial. Se recuerda que se hizo la simplificación de que los ruidos en la dinámica son independientes, es decir que el ruido está dado por una matriz normal multivariada con matriz de covarianzas diagonal. Nos interesa por tanto, solo la diagonal de la matriz, las varianzas. Se muestra el efecto en la estimación al usar distintos valores para ellas.

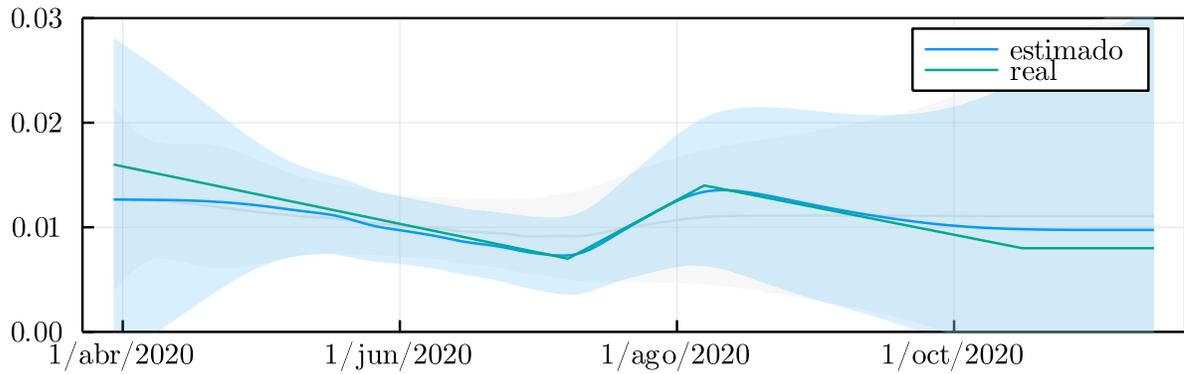


(a) Clase 1.

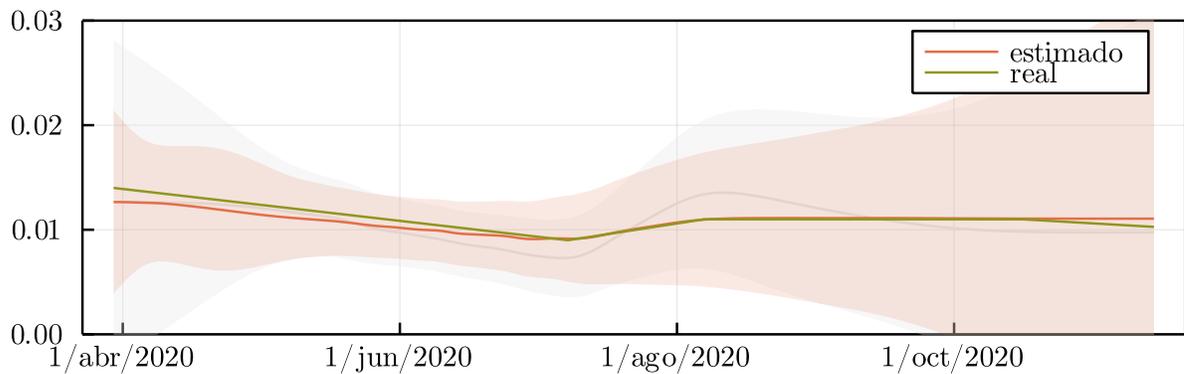


(b) Clase 2.

Figura 3.5: Número de Infectados, comparando los resultados obtenidos con la solución real, en valores absolutos y normalizados con respecto a la cantidad de personas por clase.



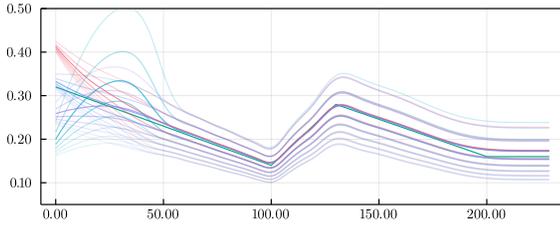
(a) Clase 1.



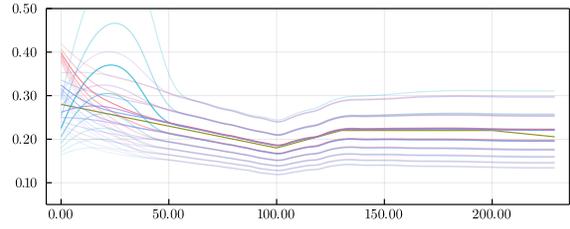
(b) Clase 2.

Figura 3.6: Factor sanitario, comparando resultados obtenidos con función de control usada para general los datos, acotando el dominio en el eje y para mejor apreciación.

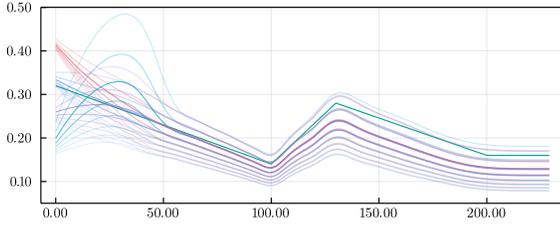
3.2.2. Sensibilidad con respecto a β_{exterior} y α_0



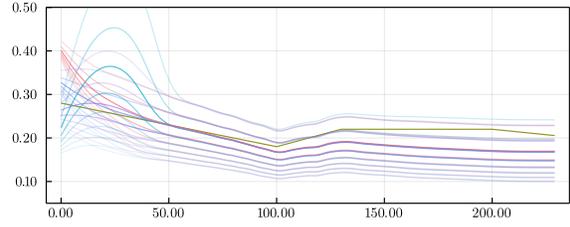
(a) Estimación control clase 1, suponiendo conocidos $\gamma_E = 1/5.1$ y $\gamma_I = 1/7.2$



(b) Estimación control clase 2, suponiendo conocidos $\gamma_E = 1/5.1$ y $\gamma_I = 1/7.2$



(c) Estimación control clase 1, con $\gamma_E = 1/5.8$ y $\gamma_I = 1/8.2$



(d) Estimación control clase 2, con $\gamma_E = 1/5.8$ y $\gamma_I = 1/8.2$



Figura 3.7: Sensibilidad ante riesgo exterior y condición inicial del factor sanitario. Cada color representa un condición inicial diferente usada en la estimación, mientras que la opacidad de la solución representa la distancia de β_{exterior} al valor real; a mayor transparencia, más lejano.

Los riesgos β y las condiciones iniciales serán valores desconocidos en la práctica. Esta sección estudia el efecto de estos valores en la solución estimada. Utilizando el caso sintético ya presentado, se estudia el efecto del valor β_{exterior} y de las condiciones iniciales para factor sanitario α en los resultados. Los datos de casos acumulados observados fueron generados con un control conocido (línea verde) y parámetros $\gamma_E = 1/5.1[\text{días}]^{-1}$, $\gamma_I = 1/7.2[\text{días}]^{-1}$ y $\beta_{\text{exterior}} = 1.88$.

Para distintos valores de β_{exterior} y distintas condiciones iniciales para el factor sanitario, se estima el factor sanitario suavizado a lo largo del tiempo y se compara con el valor de control usado para generar los datos. Los resultados pueden verse en la figura 3.7. Se estudian dos casos; uno con tasas de incubación γ_E y de remoción γ_I son conocidos (figuras 3.7a y 3.7b) y otro donde se utilizan valores pausibles para ambos parámetros (figuras 3.7c y 3.7d).

Se observa que tras una cierta cantidad de tiempo ($t \geq 50$ la condición inicial es irrelevante y todas las soluciones dadas por un mismo β_{exterior} siguen la misma trayectoria. Al comenzar en una condición inicial demasiado baja, como $\alpha_0 = 0.15$, se observa una tendencia a sobrecorregir el error antes de llegar a la trayectoria común. Para $t \geq 200$, la estimación

es menos exacta. Esto ya era visible en la figura 3.6, y probablemente se debe al hecho de que en este caso particular, para ese tiempo $S_i(t) \approx 0$, así que la solución ya no es sensible al parámetro α_i .

Cuando los parámetros γ_E y γ_I son conocidos, la estimación del factor sanitario es más certera al usar un valor β_{exterior} más cercano al real. Esto se observa en la figuras 3.7a y 3.7b, donde las estimaciones más oscuras están justo en la línea verde que marca el factor sanitario de control usado para generar los datos. En el caso en que los parámetros γ_E y γ_I son desconocidos y solo se usan valores pausibles, la estimación es menos exactata; el factor sanitario (para este caso particular) es subestimado al usar $\beta_{\text{exterior}} = 1.8$, el más cercano al real (1.88) de entre los utilizados. Esto se observa en la figuras 3.7c y 3.7d, donde las estimaciones más oscuras están por debajo de la línea verde del factor sanitario de control.

3.2.3. Sensibilidad ante la covarianza

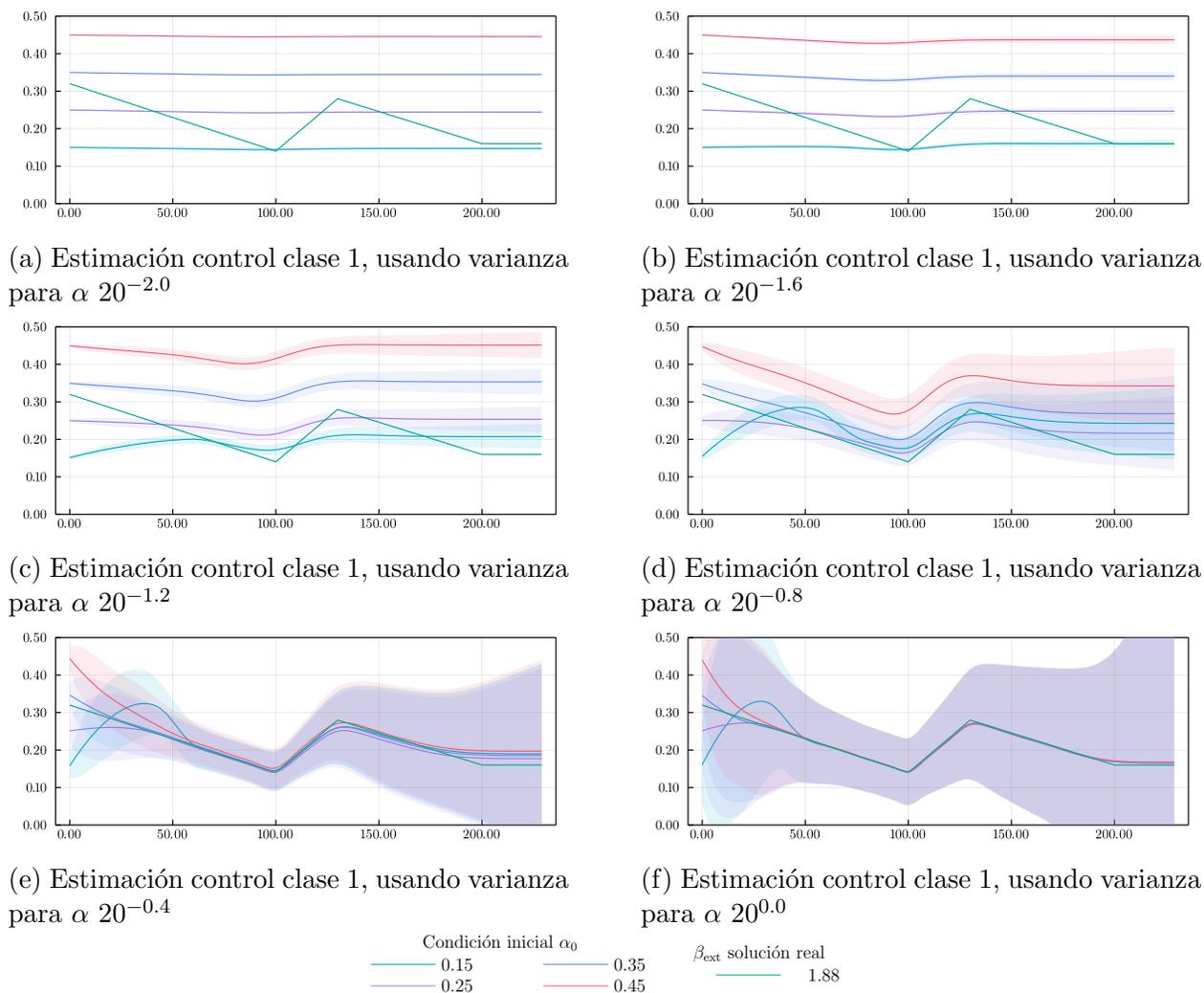


Figura 3.8: Sensibilidad ante la varianza y condición inicial del factor sanitario. Cada color representa una condición inicial diferente usada en la estimación.

Se recuerda que se hizo la simplificación de que los ruidos en la dinámica son independientes, es decir que el ruido está dado por una matriz normal multivariada con matriz de covarianzas diagonal. Nos interesa por tanto, solo la diagonal de la matriz, las varianzas. La figura 3.8 muestra la estimación del factor sanitario para distintos valores de varianza del ruido en la dinámica. Este valor refleja la certeza que tenemos en la dinámica del estado; si hay mucha, se usa una varianza baja. Si no hay tanta seguridad, se usa una varianza mayor, de tal forma de incorporar la información de los datos.

Los gráficos 3.8a y 3.8b muestran lo que ocurre al utilizar una varianza demasiado baja en el estado aumentado; la estimación es muy insensible a los datos, y se queda muy cercana a la condición inicial dada. Tanto en 3.8c como en 3.8d ya va siendo posible percibir variaciones en el control, pero aún no se ve convergencia desde las distintas condiciones iniciales. Finalmente, los gráficos 3.8e y 3.8f tienen suficiente varianza, como se ve en la convergencia de las estimaciones a partir de $t \geq 50$ aproximadamente, sin embargo 3.8f usa una varianza excesiva, ya que pueden obtenerse resultados similares con una mayor certeza.

3.2.4. Modelo con datos reales

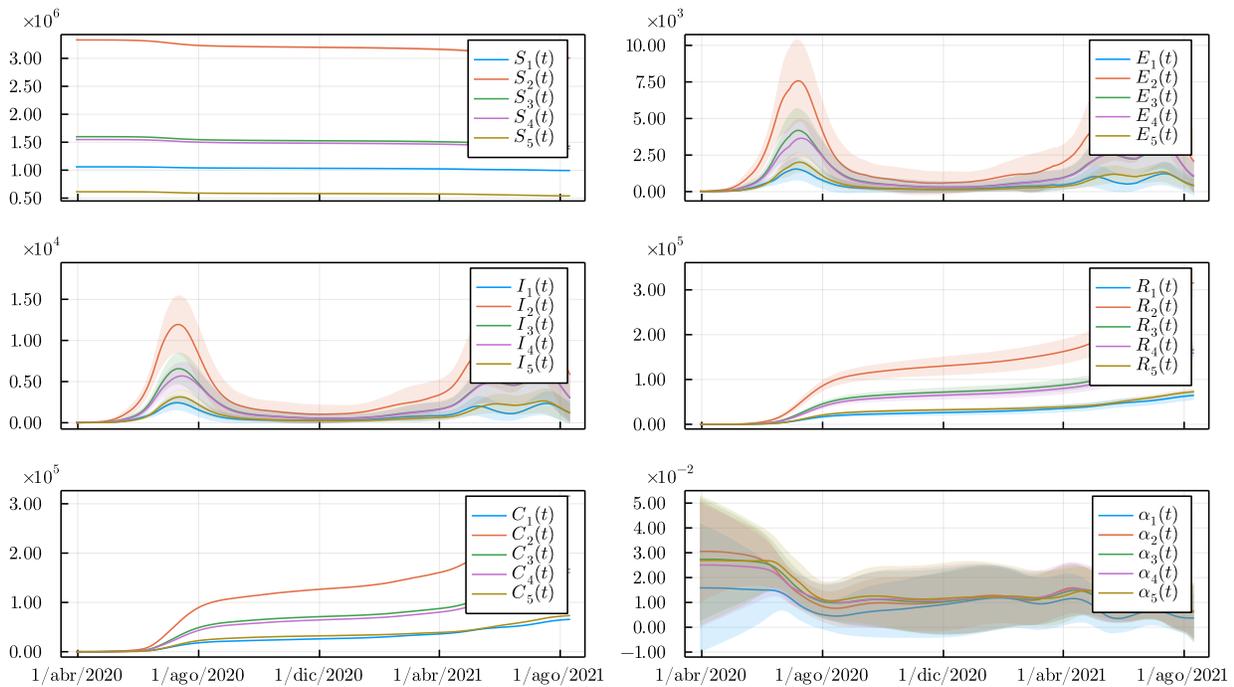


Figura 3.9: *Kalman RTS Smoother* aplicado al caso real. Se utiliza el Producto 15 de casos acumulados confirmados

El estudio realizado en las subsecciones anteriores entrega herramientas para trabajar con el caso real. Extrapolando los parámetros del filtro de Kalman utilizados en el caso sintético, es posible obtener estimaciones razonables de los estados a lo largo de la pandemia, como se ve en la figura 3.9.

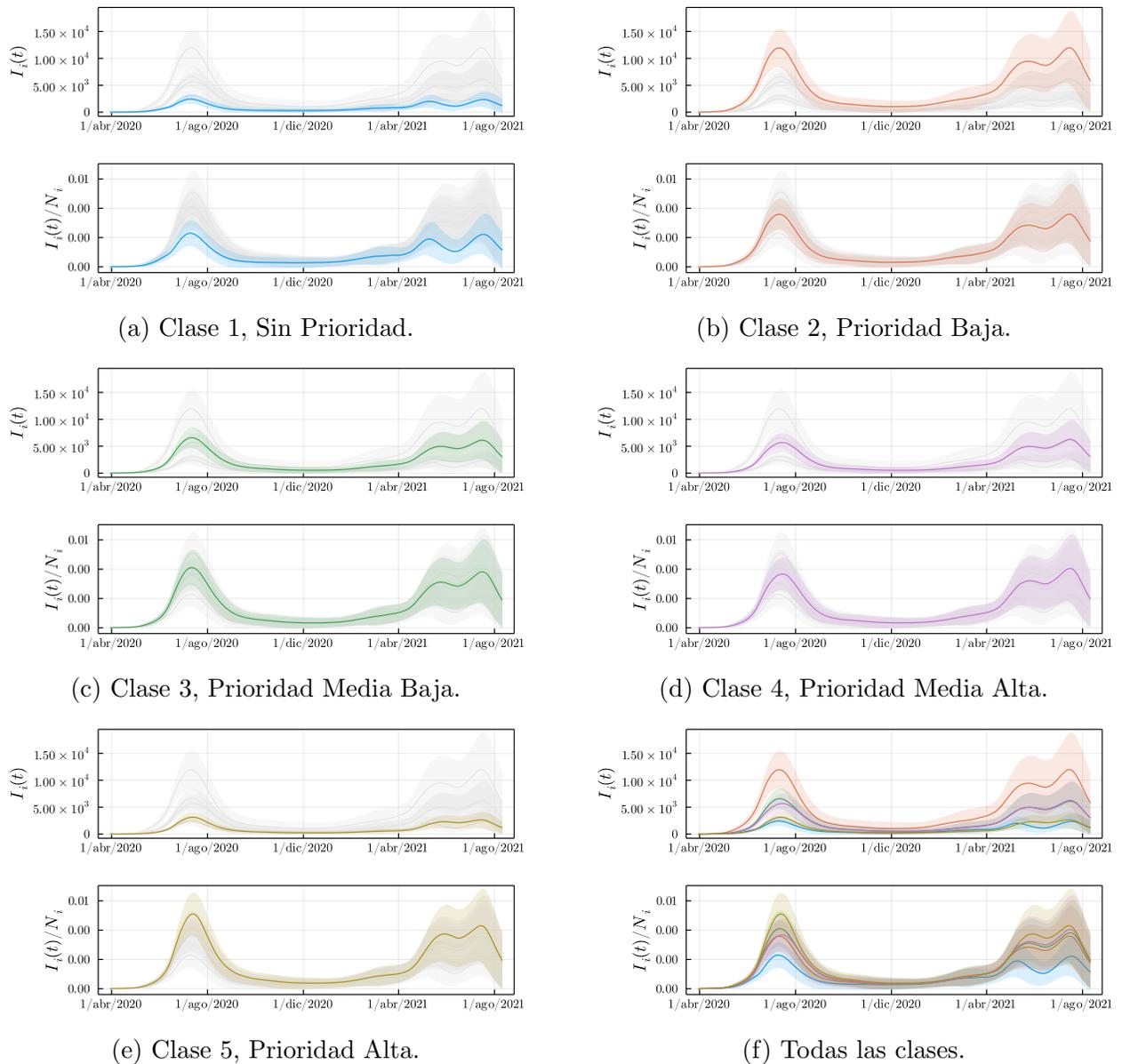


Figura 3.10: Cantidad de Infectados estimados a partir de datos reales, en valores absolutos y normalizados con respecto a la cantidad de personas por clase.

Las figuras 3.10 y 3.11 presentan en más detalle las estimaciones para el número de susceptibles e infectados respectivamente. En ambos casos se presentan tanto los valores absolutos como los normalizados con respecto al total de población de la clase. Se observa que la incidencia se encuentra ordenada de acuerdo al nivel socioeconómico; es mayor en las comunas más vulnerables y menor en las más acomodadas.

Se observa que en los meses de invierno de 2021 ocurren dos *peaks* consecutivos. Es interesante notar que cuando comienza la ola de contagios que lleva al segundo *peak*, la clase 1, la más acomodada, está en mucho mejores condiciones que a las demás; proporcionalmente alcanzó un primer *peak* más bajo, del que se recuperó con más rapidez.

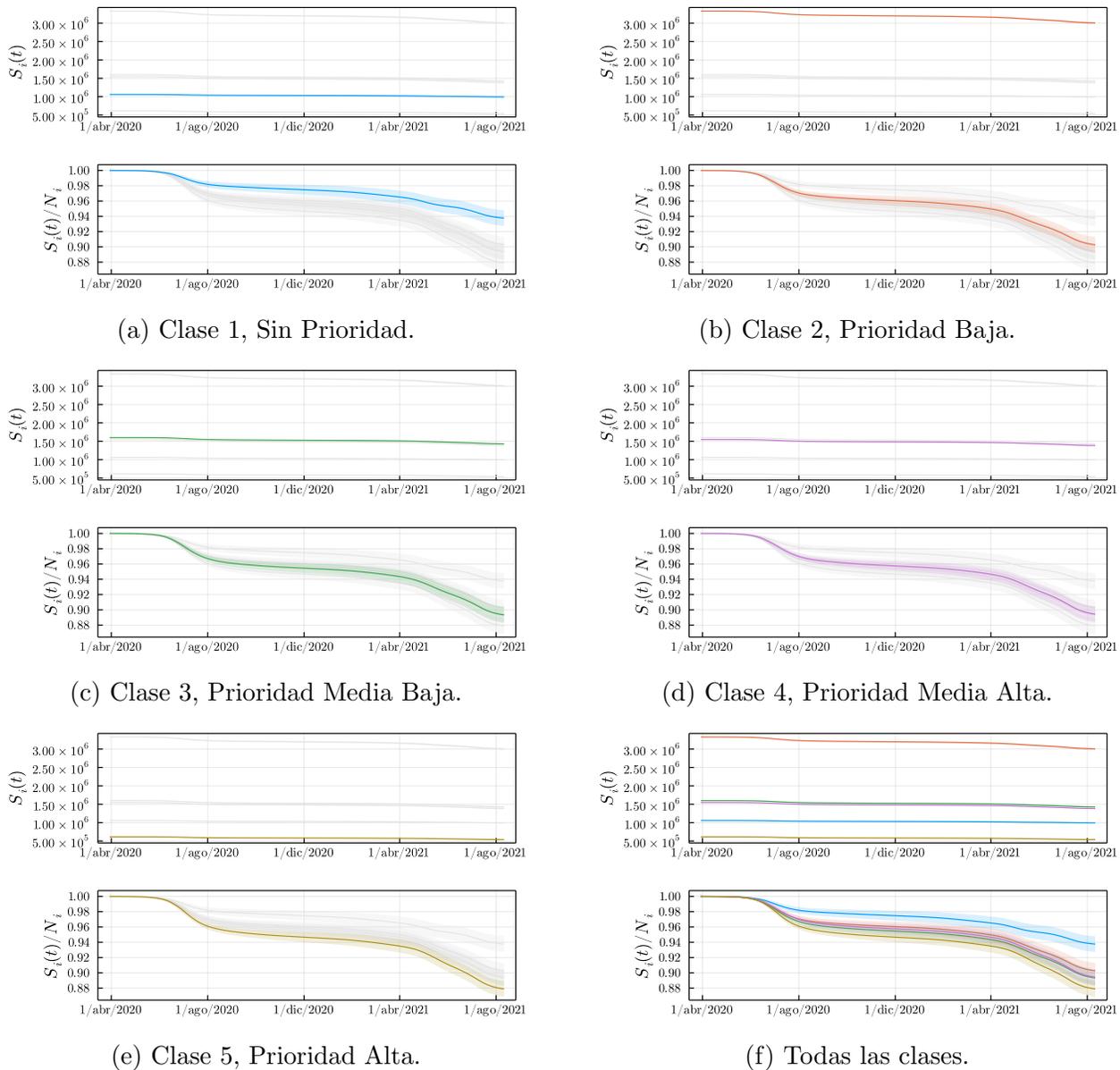
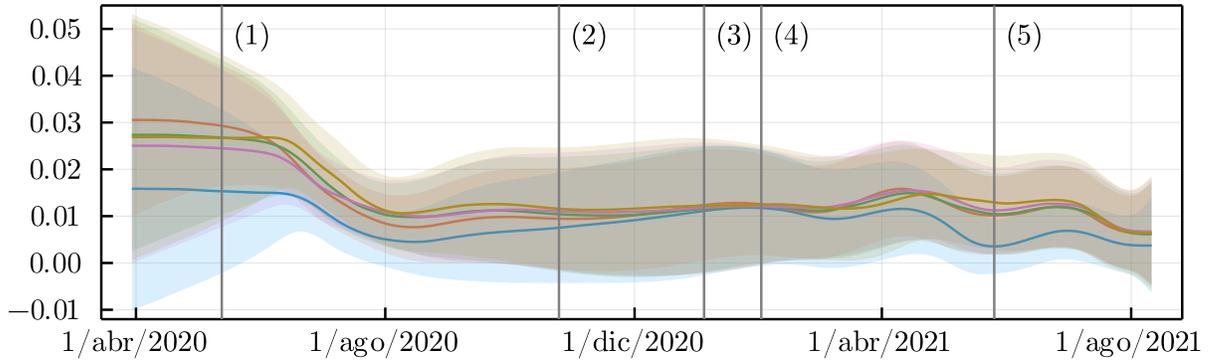


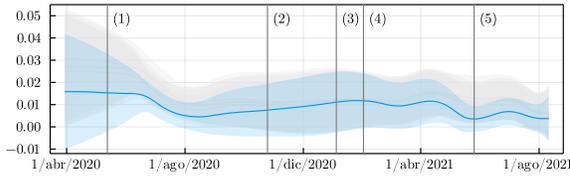
Figura 3.11: Cantidad de Susceptibles estimados a partir de datos reales, en valores absolutos y normalizados con respecto a la cantidad de personas por clase.

	Fecha	Evento
(1)	13/may/2020	Comienzo de la cuarentena total en la RM.
(2)	28/jul/2020	Comienza a regir el Plan Paso a Paso.
(3)	04/ene/2021	Comienza a funcionar el pase de vacaciones.
(4)	03/feb/2021	Comienza campaña de vacunación masiva (ver [58]).
(5)	26/may/2021	Un 50% de la población en la RM tiene la primera dosis de la vacuna.

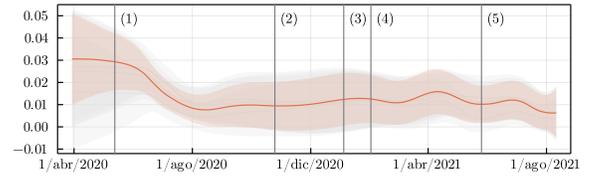
Tabla 3.5: Algunas fechas relevantes para el desarrollo de la pandemia en Santiago.



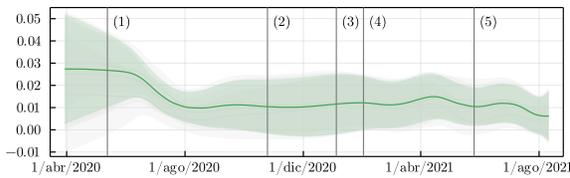
(a) Todas las clases.



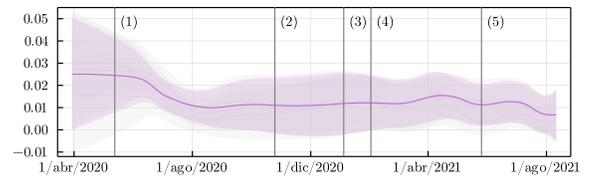
(b) Clase 1.



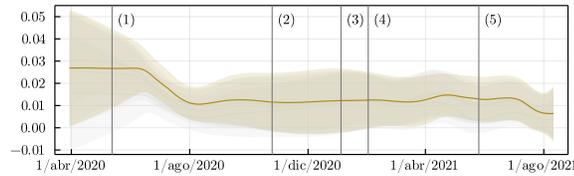
(c) Clase 2.



(d) Clase 3.



(e) Clase 4.



(f) Clase 5.

Figura 3.12: Factor sanitario estimado a partir de datos reales. Las líneas grises corresponden a las fechas relevantes de la tabla 3.5.

La tabla 3.5 muestra una lista de fechas interesantes dentro del desarrollo de la pandemia en Chile. La figura 3.12 presenta las estimaciones del factor sanitario para las distintas clases, donde se marcan estas fechas. En general, todas las clases excepto la clase 1, correspondiente a la clase Sin Prioridad o la más acomodada, se comportan de manera muy similar durante todo el período estudiado.

El factor sanitario de la clase alta, sin embargo, se mantiene consistentemente por debajo de los demás. Esta diferencia es marcada en los primeros meses de pandemia, durante el período de cuarentena, pero va desapareciendo durante los primeros meses del Plan Paso a Paso, que corresponde a la vuelta a la normalidad progresiva. La diferencia es prácticamente inexistente durante el período de vacaciones de verano de inicios de 2021. En esas fechas da inicio a la campaña de vacunación masiva, y unos meses después se observa un brusca

reducción en el factor sanitario de la clase alta, que es especialmente notoria a finales de mayo, cuando el 50% de la población de la RM ha recibido una dosis de la vacuna. Hacia finales de julio, las otras clases también experimentan una reducción importante, pero que al menos en el período estudiado no alcanza a la de la clase alta.

3.2.5. Sensibilidad ante los parámetros γ_E, γ_I

El estudio de sensibilidad mostró que los resultados dependen en mayor medida de γ_I , y que γ_E solo produce variaciones muy pequeñas en el factor sanitario. Con esto en mente, la figura 3.13 muestra la sensibilidad de los resultados con respecto a la variable γ_I . Si bien la relación entre los factores sanitario no se mantiene constante en cada caso, sí se observa claramente la brusca reducción en el factor sanitario de la clase Sin Prioridad, o la más acomodada, hacia $t = 400$, de forma mucho más marcada que las demás clases.

Si γ_I no es igual para todas las clases, entonces ya no se puede esperar que el orden entre los factores se mantenga, no tienen por qué “expandirse” o “contraerse” de la misma forma. Se observa que períodos de recuperación más largos tienden a contraer el factor sanitario y viceversa.

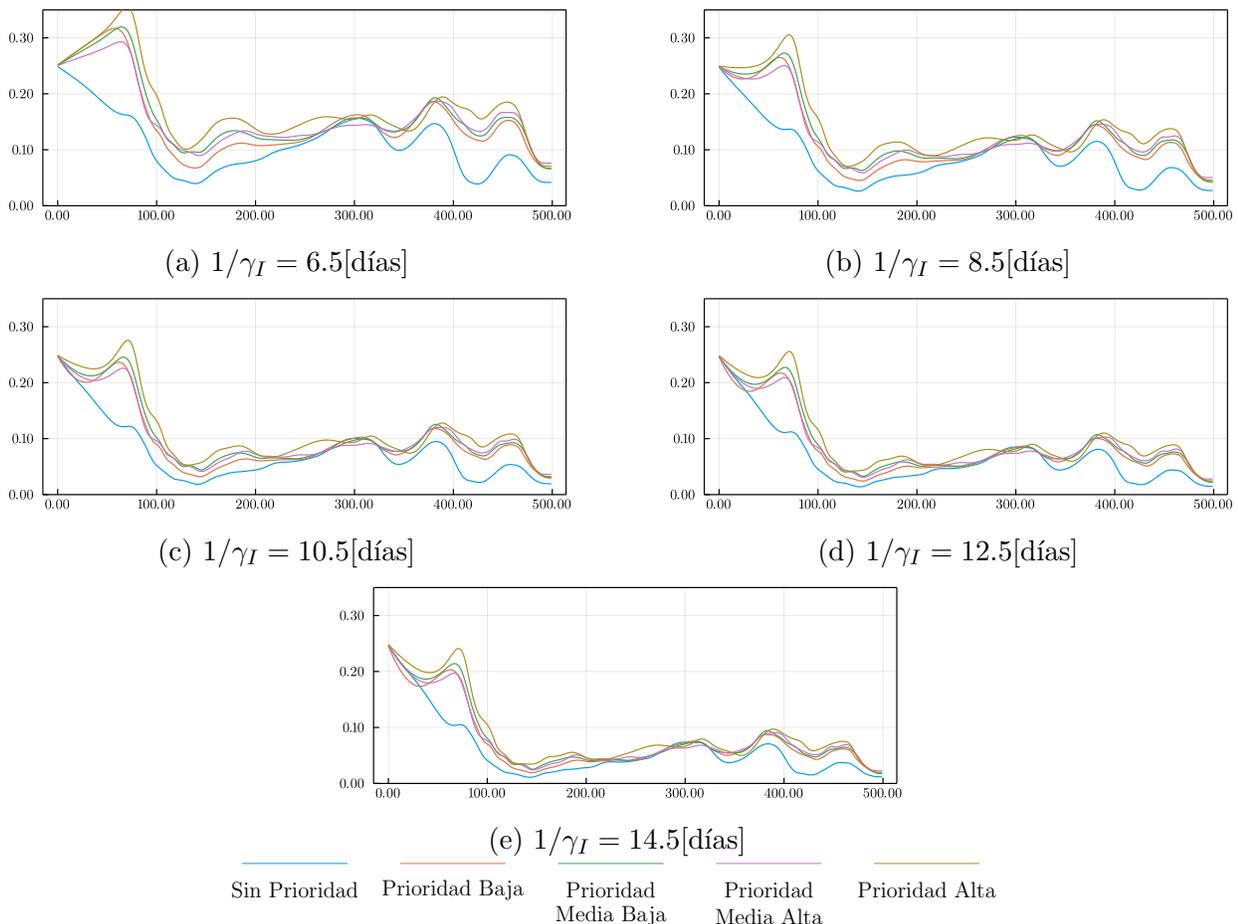


Figura 3.13

3.3. Evaluación del modelo mediante casos hipotéticos

Las combinaciones entre los distintos niveles de cuidado y cuarentena presentados en 2.5 dan lugar a una serie de casos, los cuales son presentados en la tabla 3.6. Se resuelve el sistema de ecuaciones diferenciales ordinarias 2.2 para cada caso, utilizando las distintas variantes de $\alpha_i(t)$ y $P(t)$, modificadas a partir de las estimaciones logradas en la sección anterior.

	Cuidado insuficiente	Cuidado normal	Cuidado extra
Sin cuarentena	Caso 1	Caso 2	Caso 3
Cuarentena normal	Caso 4	Caso 5	Caso 6
Cuarentena fuerte	Caso 7	Caso 8	Caso 9

Tabla 3.6: Casos hipotéticos para distintas combinaciones de cuarentena y cuidado.

Se prueban dos escenarios. En el primero de ellos, β_{exterior} es un valor mucho más grande que 1, de forma que la tasa de contagio está totalmente dominada por el ambiente exterior y prácticamente todos los contagios ocurren fuera de casa. Este se estudia en la subsección 3.3.1. Este escenario, como se mencionó en la sección 2.3.1, no es consistente con la realidad, puesto que aproximadamente un tercio de los casos se contagia en el hogar [65]. En el segundo escenario, β_{exterior} es un valor mayor pero cercano a 1, de tal forma que el ambiente hogar contribuye a la tasa de contagios, lo que es más realista. Este se estudia en 3.3.2.

3.3.1. Escenario con contagio exterior predominante

En este escenario se considera $\beta_{\text{exterior}} \gg 1$, de forma que la tasa de contagio está totalmente dominada por el ambiente exterior y las infecciones ocurren solamente en el exterior del hogar, las infecciones dentro del hogar son despreciables en comparación. Se utiliza $\beta_{\text{exterior}} = 68.0$, pero los resultados son similares para un rango amplio de parámetros.

Como era de esperarse, las diferencias socioeconómicas en la incidencia disminuyen debido a que la mayoría de los casos están utilizando, o bien la misma matriz de tiempos de residencia, o bien los mismos factores sanitarios, o incluso ambos.

La tabla 3.7 describe los resultados para los distintos casos considerados, tomando como referencia el caso 5, de cuarentena y cuidado normal. En los casos 1 y 2, que corresponden a casos sin cuarentena con cuidado insuficiente o cuidado normal respectivamente, se contagia la gran mayoría de la población dentro de los primeros 6 meses considerados. En el caso con cuidado normal, la clase más acomodada resulta tener una incidencia bastante menor que las demás.

El cuidado extra logra contener la enfermedad con cuarentena normal o fuerte, como muestran los casos 6 y 9. La enfermedad desaparece rápidamente de entre la población en los primeros meses, aunque esto desde luego no considera importación de casos externos.

	Cuidado insuficiente	Cuidado normal	Cuidado extra
Sin cuarentena	<i>Caso 1:</i> Se contagia un 90 % de la población en los primeros 6 meses, de manera uniforme entre las distintas clases. Luego la enfermedad desaparece.	<i>Caso 2:</i> Se contagia un 65 % de la población de la clase alta en los primeros 6 meses, y un 80 % de las demás clases. Luego la enfermedad desaparece.	<i>Caso 3:</i> Se logra controlar el alza de casos, alcanzando el <i>peak</i> de mediados de julio de 2020 y el de inicios de mayo de 2021 con menos de la mitad de casos, e incluso se evita el rebrote de junio de 2021. Ver figura 3.14a.
Cuarentena normal	<i>Caso 4:</i> El primer <i>peak</i> de julio de 2020 alcanza más del doble de casos y tiene una reducción lenta, de tal forma que el <i>peak</i> de mayo de 2021 es apenas registrado como un alza ligera. Ver figura 3.14b.	<i>Caso 5:</i> Desarrollo real de la pandemia, tomado como referencia	<i>Caso 6:</i> La enfermedad alcanza un pequeño <i>peak</i> en julio de 2020 y luego se extingue.
Cuarentena fuerte	<i>Caso 7:</i> El primer <i>peak</i> es muy similar en fecha y número de infectados, pero con un alza sostenida en el número de casos a partir septiembre de 2020, llegando a un <i>peak</i> de más del doble de infectados en julio de 2021. Ver figura 3.14c.	<i>Caso 8:</i> Se observan dos <i>peaks</i> de unas 5000 personas, similar al Caso 3 pero sin poder contener el rebrote de junio de 2021. Ver figura 3.14d.	<i>Caso 9:</i> La enfermedad se extingue luego de los primeros meses.

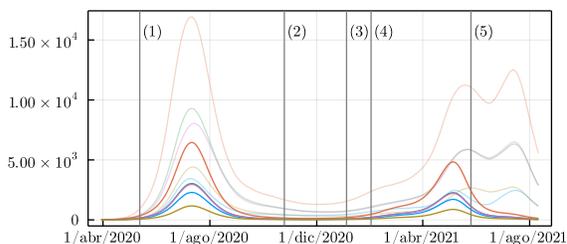
Tabla 3.7: Observaciones para cada caso hipotético, escenario con $\beta_{\text{exterior}} = 68$.

La figura 3.14 muestra cuatro casos interesantes, comparándolos con la situación normal. En estos tres casos, el impacto es mayor sobre la clase 5 (prioridad alta, la más vulnerable). Esto puede atribuirse al hecho de que esa es la clase que presenta en general un mayor factor sanitario y movilidad, por lo que se verían más favorecidos si pudieran guardar las cuarentenas o las medidas de cuidado de mejor manera.

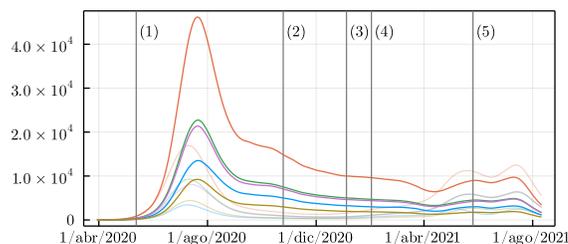
El caso 3, que puede verse en 3.14a, es interesante, ya que propone que la pandemia pudo controlarse incluso sin cuarentenas ni reducciones de movilidad, siempre y cuando todos pudieran guardar adecuadamente las distintas medidas de cuidado, evitando aglomeraciones, usando mascarilla, practicando el lavado de manos frecuente, etc. El caso 4, en la figura 3.14b,

coincide en que las medidas de higiene y cuidado fueron importantes, y plantea que evitaron un *peak* de más del doble de casos en los primeros meses de la pandemia.

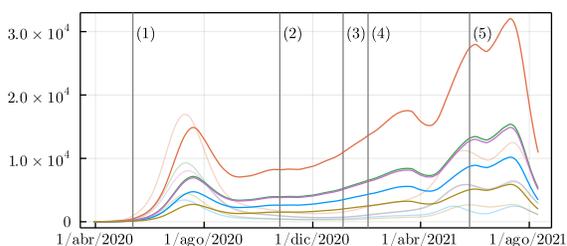
El caso 7, en la figura 3.14c, afirma que una cuarentena más fuerte habría compensado una falta de preocupación o capacidad de cumplimiento de las medidas de higiene y cuidado, al menos en los primeros meses. La situación se desborda a medida que se acerca el 2021, debido a que la población sale del confinamiento. El caso 8, en la figura 3.14d, es muy similar al caso 3, pero no logra evitar un segundo *peak* en julio de 2021.



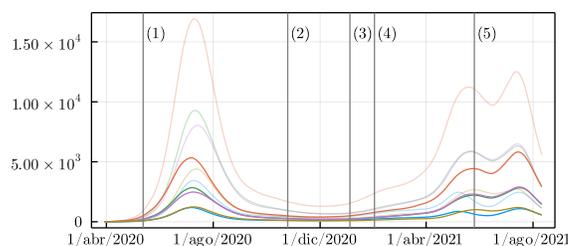
(a) Caso 3: Cuidado extra sin cuarentena.



(b) Caso 4: Cuidado insuficiente y cuarentena normal.



(c) Caso 7: Cuidado insuficiente y cuarentena fuerte.



(d) Caso 8: Cuidado normal y cuarentena fuerte.

— Sin Prioridad — Prioridad Baja — Prioridad Media Baja — Prioridad Media Alta — Prioridad Alta

Figura 3.14: Personas infectadas para caso hipotéticos seleccionados, con la estimación de los casos reales de fondo. Diferentes límites para el eje y . Las líneas grises corresponden a las fechas relevantes de la tabla 3.5.

3.3.2. Escenario con contagio en hogar

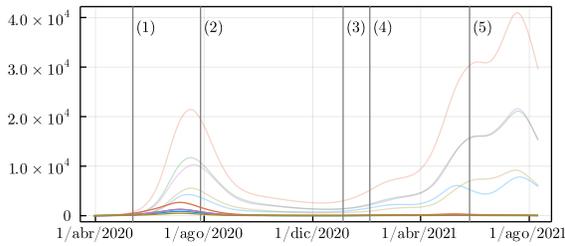
En este caso se buscó un valor donde existiera tanto contagio en el hogar como en el exterior; esto se hizo estudiando las contribuciones de los ambientes a la tasa de contagio, calculando las contribuciones a la derivada de cada ambiente dadas por 2.25, en la subsección 2.4.3, y verificando que las del hogar no fueran despreciables en comparación a las del exterior. El valor elegido fue $\beta_{\text{exterior}} = 1.2$.

Bajo estas condiciones, se observa que la cuarentena prácticamente no influye en el número de contagios; lo que realmente determina si se controla o no la pandemia es el factor sanitario. Los resultados se sintetizan en la tabla 3.8, y los casos de interés elegidos anteriormente se presentan en la figura 3.15. Con cuidado extra es posible controlar completamente la

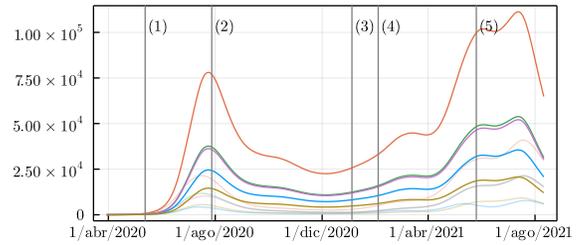
pandemia, incluso sin cuarentena. Este modelo, sin embargo, dice que la cuarentena no tuvo ningún impacto, y que bastaban las medidas de cuidado. Finalmente, este modelo predice que si toda la población se hubiera visto enfrentada al mismo riesgo que la clase más vulnerable, la pandemia habría sido muchísimo más grave, alcanzando *peaks* de más del doble de casos.

	Cuidado insuficiente	Cuidado normal	Cuidado extra
Con o sin cuarentena	<i>Casos 1, 4 y 7:</i> Se observan <i>peaks</i> de más del doble de casos que el normal, y que se reducen más lentamente. El número de contagios no baja del 0.5% de la población. Ver figuras 3.15b y 3.15c.	<i>Caso 2 y 8</i> Muy similares al caso 5, el normal usado como referencia. Ver figura 3.15d.	<i>Caso 3, 6, y 9:</i> La enfermedad tiene un pequeño <i>peak</i> en torno a julio de 2020 y se extingue luego de unos meses. Ver figura 3.15a.

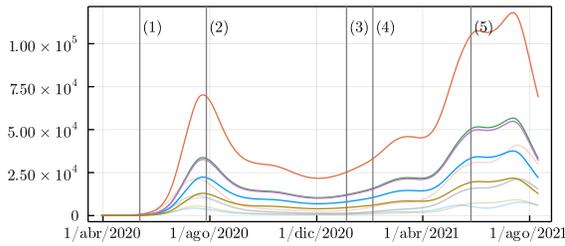
Tabla 3.8: Observaciones para cada caso hipotético, escenario con $\beta_{\text{exterior}} = 1.2$. Tabla condensada, puesto que no hay variaciones importantes con respecto a la fuerza de la cuarentena.



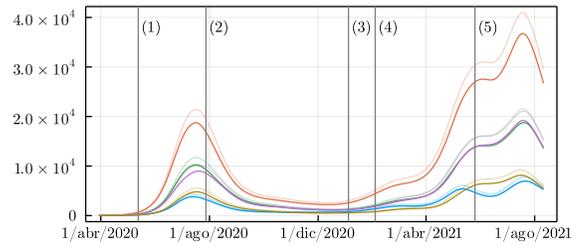
(a) Caso 3: Cuidado extra sin cuarentena.



(b) Caso 4: Cuidado insuficiente y cuarentena normal.



(c) Caso 7: Cuidado insuficiente y cuarentena fuerte.



(d) Caso 8: Cuidado normal y cuarentena fuerte.

— Sin Prioridad — Prioridad Baja — Prioridad Media Baja — Prioridad Media Alta — Prioridad Alta

Figura 3.15: Personas infectadas para caso hipotéticos seleccionados, con la estimación de los casos reales de fondo. Diferentes límites para el eje y . Las líneas grises corresponden a las fechas relevantes de la tabla 3.5.

Capítulo 4

Discusión

A continuación se ofrece una discusión con respecto al trabajo realizado. La sección 4.1 comenta los resultados obtenidos para el caso de estudio y para las situaciones hipotéticas, resaltando las limitaciones derivadas de los supuestos y simplificaciones hechos a lo largo del modelamiento, y presentando posibles interpretaciones. La sección 4.2 comenta la metodología seguida, presenta algunos enfoques alternativos con que se podría haber abordado el trabajo, expone posibles extensiones y discute el cumplimiento de los objetivos.

4.1. Acerca del caso de estudio

En primer lugar, es necesario recordar los supuestos y simplificaciones hechos a la hora de modelar el caso de estudio. Por un lado, se consideró que la población en la Región Metropolitana se mantenía constante, despreciando efectos como la natalidad, mortalidad debido a causas naturales y efectos migratorios entre ciudades.

Por otro lado, algunos factores relevantes a la transmisión del COVID-19 fueron dejados fuera al utilizar un modelo SEIR, como la existencia de casos sin síntomas o con síntomas leves, que pueden o no ser reportados y que también transmiten el virus [52][53][54]. Tampoco fueron considerados los efectos del contagio en la movilidad; en general las personas con síntomas son detectadas y puestas en cuarentena. En este modelo sin embargo se siguen comportando con normalidad. El *framework* [18] considera esta situación, y la resuelve usando matrices de tiempos de residencia distintas para los contagiados.

Una medida de gran relevancia que no fue considerada en el modelo fue la vacunación. Esta, sin embargo, debería verse directamente reflejada en una disminución importante del factor sanitario una vez que se vacuna gran parte de la población. El supuesto de que todas las clases socioeconómicas tienen la misma tasa de incubación y recuperación es un factor que debe ser considerado, especialmente al tener en cuenta el estudio de sensibilidad presentado en 3.2.5.

El uso del filtro de Kalman tiene, desde luego, limitaciones; teóricamente solo puede justificarse en el caso lineal, y el uso del filtro extendido lleva a errores numéricos que podrían

mitigarse usando un filtro más apto para tratar con no linealidades, como el filtro *unscented*. El estudio del caso sintético muestra también que el filtro tiene dificultades para estimar los valores sanitarios al comienzo de la pandemia, cuando hay muy pocos casos, por lo que los resultados en la fase inicial son poco confiables. Los estudios de sensibilidad mostraron además las diferencias en las estimaciones, dependiendo de los valores de varianzas elegidas como parámetros del filtro.

El número de casos detectados no es la mejor observación; se ha hablado de cómo la capacidad de detectar los nuevos contagios difiere entre clases socioeconómicas [32], sugiriendo que el número de fallecidos es una mejor medida. Esto sin embargo, sugiere que el número de casos está subestimado para las clases más vulnerables, lo que solo contribuye a resaltar los resultados obtenidos.

Con respecto a las estimaciones calculadas, se destaca el hecho de que el factor sanitario de la clase más acomodada se mantiene por debajo de las demás clases. Ese resultado es robusto ante los análisis de sensibilidad. Se intentó inicialmente obtener un factor sanitario común a todas las clases, lo que sugeriría que todos enfrentan el mismo riesgo en el exterior, pero no fue posible obtener resultados razonables.

Todo esto apunta a que las diferencias en movilidad por sí solas no son suficientes para explicar las diferencias en la cantidad de contagios. Las personas pertenecientes a los niveles socioeconómicos más bajos no solo no pueden guardar las cuarentenas de la misma forma que las clases acomodadas, sino que además, al salir al exterior, se encuentran en ambientes más riesgosos, ya sea por estar en ambientes más concurridos, o por no utilizar adecuadamente las medidas de seguridad como el lavado de manos y el uso correcto de mascarillas. Estos resultados son consistentes con los encontrados por [15]. Otras posibles explicaciones podrían ser las diferencias en la capacidad de testeo y trazabilidad [33][32].

Con respecto al ajuste del factor sanitario, a pesar de la gran incerteza que surge de utilizar valores grandes de covarianza, el caso sintético sugiere que se obtienen mejores resultados de esta forma. Los resultados deben considerar además los estudios de sensibilidad de la subsección 3.2.5. Mientras la hipótesis $\gamma_{Ei} = \gamma_E$ y $\gamma_{Ii} = \gamma_I$ sea válida, es decir, que las tasas de incubación y remoción son iguales en cada clase, se puede esperar que las estimaciones del factor sanitario α_i se expandan o contraigan uniformemente. La tasa de exposición γ_E no tiene gran relevancia, sin embargo, la de recuperación γ_I puede hacer variar considerablemente los resultados. Ahora bien, para subestimar el factor sanitario, se debería tener un tiempo de recuperación mayor, pero esto es improbable considerando que el acceso a salud es superior en las clases más acomodadas.

Destaca el factor sanitario en el periodo mayo-junio de 2021, donde se observa que el grupo más acomodado consigue una reducción mucho más marcada que las demás. Una posible hipótesis sería atribuir esto al avance del proceso de vacunación en Chile, que para esas fechas ya había superado el 50 % de población vacunada en la Región Metropolitana (ver figura 2.4); ¿puede atribuirse la reducción del factor sanitario a la campaña de vacunación? y de ser así ¿por qué el proceso parece aliviar primero a las comunas acomodadas? Una posible hipótesis es atribuir la diferencia a una mayor disposición de ese grupo a recibir la vacuna. Otra posibilidad es considerar la distribución de adultos mayores, que tienen una mayor proporción en las comunas acomodadas mayor [81], recordando que la campaña de

vacunación ocurrió por tramos de edad, de mayores a jóvenes.

Con respecto a los casos hipotéticos, se observan dos escenarios que dependen del valor del riesgo β_{exterior} . Para este análisis es necesario considerar el aporte a la derivada a cada uno de los ambientes, calculado en 2.4.3. Por un lado, al usar un valor notoriamente mayor que 1, el aporte a la derivada dado por el ambiente **hogar** es despreciable en comparación al dado por el **exterior**. Esto quiere decir que prácticamente todos los contagios ocurren cuando la gente sale de su hogar. Esto, sin embargo, no es consistente con [65][19], que afirman que aproximadamente 1/3 de los contagios ocurren en el hogar.

En esta situación se observa que las medidas de cuidado son sumamente relevantes; los casos “Cuidado extra sin cuarentena” y “Cuidado normal con cuarentena fuerte” son muy similares, salvo que las mejores medidas de cuidado logran prevenir un segundo *peak* en agosto de 2021. En los otros casos de “Cuidado extra”, con cuarentena normal o fuerte, se logra controlar la pandemia dentro de los primeros seis meses.

Por otro lado, cuando se usa un valor ligeramente mayor a 1, el aporte del ambiente **hogar** a la derivada ya es perceptible y comparable con el de **exterior**. El estudio de los casos hipotéticos, sin embargo, muestra que la cuarentena casi no influya en nada en los resultados finales, y que todo depende de las medidas de cuidado.

El modelo en ambos escenarios coinciden en que si fuera posible que todos se cuidaran al mismo nivel que el grupo más acomodado, se podría controlar el virus incluso sin la exigencia de cuarentenas.

4.2. Acerca de la metodología

Una de las diferencias distintivas de este enfoque es utilizar los datos de movilidad para estimar tiempos de residencia, siguiendo la forma lagrangiana. Esta no era la única forma posible, pues los datos de la Encuesta Origen-Destino contienen información detallada de viajes, incluyendo comunas de salida y llegada. Una forma alternativa de usar la Encuesta habría sido utilizar sus datos como movilidad base entre las distintas comunas, y aprovechar los datos de movilidad en pandemia, como los del ISCI presentados en la sección 2.1.5, para modificar la cantidad de viajes, siguiendo una metodología similar a [11], aunque debe recordarse que los datos tienen ya una década, y podrían estar bastante desactualizados.

De seguir esta forma de trabajo, habría sido posible definir algo análogo al factor sanitario, para cada zona/comuna. El no tener una distinción entre clases y ambientes sin embargo, impone el tener que trabajar únicamente con divisiones espaciales, como zonas censales, comunas y agrupaciones de comunas, dificultando considerar otras clases dadas por grupos etarios, sexo, etc.

Ahora, si bien la metodología propuesta sí permite trabajar simultáneamente con clases y dispersión espacial, las dificultades no se eliminan completamente, puesto que sigue siendo necesario estimar los tiempos de residencia para cada una de las clases, y la forma en que estos variaron en cada ambiente no fue uniforme en distintos grupos etarios.

A modo de ejemplo, podría ser posible considerar grupos etarios, puesto que existen series de tiempo de número de contagios agrupados por edad. La variación de movilidad por grupo etario, sin embargo, no es tan fácil de obtener, considerando que los datos de movilidad no tienen ese nivel de granularidad, y aunque lo tuvieran, suelen usar fuentes que dejan mal representados ciertos grupos; los datos de ISCI por ejemplo, se derivan del uso de torres de telefonía móvil, que en el caso de niños y adultos mayores podría no ser representativo.

Al intentar agregar más ambientes al análisis, surgen problemas similares. Los datos más cercanos a movilidad por ambiente en pandemia son los provistos por Google, presentados en la sección 2.1.4. Estos datos están agrupados a nivel regional/provincial, pero se sabe [31] que no son uniformes entre distintos niveles socioeconómicos; el teletrabajo es mucho más común en niveles altos, mientras que un porcentaje importante de la población más vulnerable tiene trabajos esenciales que no pueden detenerse en pandemia.

Debido a la dificultad de actualizar la matriz de tiempos de residencia de una forma que no hiciera un gran cantidad de supuestos, la Encuesta Origen-Destino terminó jugando un rol muy secundario en el trabajo, ya que su utilidad se vio relegada a una simple estimación del tiempo que pasaban en el hogar, antes de la pandemia, los distintos grupos socioeconómicos. Debe recordarse que esta estimación se hizo con datos del 2012, podría ser bastante inexacta. Los valores obtenidos, sin embargo, sí son interesantes, ya que el mayor tiempo de la clase acomodada en el exterior en condiciones normales “amortigua” su mayor reducción de movilidad, haciendo que la diferencia de tiempos en el exterior sea más reducida de lo que las variaciones de movilidad como 1 dan a entender. Pero no se puede negar que la EOD no se usó tanto como se deseaba en un principio, y que habría sido aprovechada de mejor manera siguiendo una metodología como la euleriana antes propuesta.

Por otro lado, la metodología desarrollada en este trabajo, presenta la ventaja de ser utilizable con menos datos de movilidad disponibles, por lo que es aplicable incluso en regiones donde no se tenga el detalle de las tasas de movilidad entre cada par de zonas; solo es necesario saber cuánto se mueve la gente, no hacia dónde lo hace, debido a que todos interactúan en un ambiente **exterior** común.

Una desventaja no prevista en un principio es la estimación de los riesgos específicos por ambiente; no es tan clara como se supuso. Esto quedó en evidencia al observar los resultados de la sección 3.3. Estimaciones de la forma: “aproximadamente 1/3 de los contagios ocurren en el hogar” [65] no se traducen directamente en asignar $2\beta_{\text{hogar}} = \beta_{\text{exterior}}$.

Sería deseable aplicar la metodología a una mayor variedad de ambientes, como trabajo, transporte público, compras, etc. Incorporar la vacunación al estudio también es interesante, así como considerar compartimientos como fallecidos, hospitalizados, etc. También sería interesante utilizar otras variantes del filtro de Kalman, como el filtro H^∞ que es más confiable ante incertezas del modelo, o el *unscented*, que funciona mejor con modelo no lineales.

Conclusión

Los objetivos específicos fueron logrados en su mayor parte, puesto que fue posible plantear una forma de estimar el factor sanitario y realizar una implementación para el caso de estudio. Sin embargo, era deseable poder incluir en el modelo ambientes como **trabajo**, **estudios**, **compras** o **transporte público**, o considerar grupos etarios, y esto no fue logrado, algo que podría considerarse usando los datos de Google vistos en 2.1.4.

Con respecto al caso de estudio, los resultados sugieren que las comunas de menores ingresos no solo pudieron disminuir menos su movilidad que las comunas de mayores ingresos, sino que también se vieron expuestas a ambientes más riesgosos, en concordancia con [15]. Se observó además una brusca reducción en el factor sanitario estimado a finales de mayo de 2021, especialmente para el nivel socioeconómico más alto. Los resultados obtenidos de los casos hipotéticos subrayan en mayor o menor medida la importancia de las medidas de cuidado e higiene.

Con respecto al *framework* lagrangiano puede decirse que no aprovechó los datos de la Encuesta Origen-Destino, algo que sí podría haberse hecho con un modelo Euleriano como el del [11]. En contraste, la ventaja del modelo aplicado está en su sencillez, ya que solo requiere de datos de movilidad por clase, sin requerir los orígenes o destinos de los viajes. Los riesgos no son tan intuitivos de estimar como se pensó en primera instancia.

Las principales contribuciones de este trabajo son las siguientes. En primer lugar, la propuesta de una metodología que permite estudiar las medidas de cuidado de forma independiente a la movilidad. En segundo lugar, la aplicación del *framework* lagrangiano de clases y ambientes presentado por [18] a un caso de estudio, lo cual no había sido hecho con anterioridad. Este trabajo guarda, sin embargo, cierta similitud con [38], que plantea un modelo que incorpora patrones de actividad.

En tercer lugar, la implementación en Matlab de la técnica propuesta por [24] de obtención de uso del tiempo partir de los viajes de una Encuesta Origen-Destino. Esta implementación es de código abierto y se encuentra disponible en el repositorio <https://github.com/tabitaCatalan/lagrangian-time>. Los resultados de uso del tiempo de la Encuesta Origen-Destino Santiago 2012 son también interesantes.

En cuarto lugar, el desarrollo de una librería, escrita en el lenguaje Julia, para el trabajo con Filtro de Kalman con dinámica y observadores lineales y/o no lineales. Esta se encuentra documentada, es modular, extensible y de código abierto. Está disponible en el repositorio <https://github.com/tabitaCatalan/kalman>. Existen librerías similares en el mismo len-

guaje como `KalmanFilters.jl` y `Kalman.jl`, pero estas no fueron usadas debido a que en la etapa de decidir qué Filtro utilizar, ninguna contaba con todas las opciones que se quería explorar.

En quinto lugar, el planteamiento de una metodología para el ajuste de los hiperparámetros relacionados a la covarianza del filtro de Kalman, al estimar parámetros de un modelo epidemiológico. Este era un vacío en trabajos como [29] y [25], donde se limitan a elegir un valor fijo para el caso particular que trabajan, sin ofrecer una justificación.

Finalmente, la aplicación del modelo al desarrollo de la pandemia de COVID-19 en la ciudad de Santiago de Chile, lo que permite una estimación de la vulnerabilidad/factor sanitario, de forma independiente a la movilidad. Modelos anteriores de Santiago como [34] solo consideran movilidad.

Además, el repositorio <https://github.com/tabitaCatalan/CovidMTK> contiene el código utilizado en el análisis del caso de estudio del COVID-19 en Santiago. Este repositorio hace uso de la novedosa librería `ModelingToolkit.jl` [82], aun en una fase inicial de desarrollo, la cual, de forma similar a *software* como Modelica o Simulink, ofrece herramientas para el trabajo con modelos de ecuaciones diferenciales, cálculo simbólico de derivadas, de forma eficiente, y extensible.

Algunas posibles extensiones del trabajo realizado incluyen aplicar la metodología propuesta a otros modelos, compartimientos, parámetros y observaciones. Sería interesante considerar los efectos de la vacunación en los compartimientos, usar comunas en lugar de grupos socioeconómicos, usar un filtro que se comporte mejor frente a no linealidades (como el filtro *unscented* o “sin olor”), ampliar el modelo para considerar fallecidos, entre otros. Además, sería interesante estudiar los efectos de la vacunación considerando los distintos niveles socioeconómicos.

Bibliografía

- [1] Secretaria Regional Ministerial de Desarrollo Social y Familia de la Region Metropolitana de Santiago, “Índice De Prioridad Social de Comunas de la Región Metropolitana de Santiago 2019,” pp. 1–22, 2019.
- [2] E. Dong, H. Du, and L. Gardner, “An interactive web-based dashboard to track COVID-19 in real time,” *The Lancet Infectious Diseases*, vol. 20, pp. 533–534, may 2020.
- [3] T. Greenhalgh, J. L. Jimenez, K. A. Prather, Z. Tufekci, D. Fisman, and R. Schooley, “Ten scientific reasons in support of airborne transmission of SARS-CoV-2,” *The Lancet*, vol. 397, pp. 1603–1605, may 2021.
- [4] S. Flaxman, S. Mishra, A. Gandy, H. J. T. Unwin, T. A. Mellan, H. Coupland, C. Whitaker, H. Zhu, T. Berah, J. W. Eaton, M. Monod, P. N. Perez-Guzman, N. Schmit, L. Cilloni, K. E. C. Ainslie, M. Baguelin, A. Boonyasiri, O. Boyd, L. Cattarino, L. V. Cooper, Z. Cucunubá, G. Cuomo-Dannenburg, A. Dighe, B. Djaafara, I. Dorigatti, S. L. van Elsland, R. G. FitzJohn, K. A. M. Gaythorpe, L. Geidelberg, N. C. Grassly, W. D. Green, T. Hallett, A. Hamlet, W. Hinsley, B. Jeffrey, E. Knock, D. J. Laydon, G. Nedjati-Gilani, P. Nouvellet, K. V. Parag, I. Siveroni, H. A. Thompson, R. Verity, E. Volz, C. E. Walters, H. Wang, Y. Wang, O. J. Watson, P. Winskill, X. Xi, P. G. T. Walker, A. C. Ghani, C. A. Donnelly, S. Riley, M. A. C. Vollmer, N. M. Ferguson, L. C. Okell, and S. Bhatt, “Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe,” *Nature*, vol. 584, pp. 257–261, aug 2020.
- [5] C. Castillo-Laborde, T. de Wolff, P. Gajardo, R. Lecaros, G. Olivar-Tost, and H. Ramírez, “Assessment of event-triggered policies of nonpharmaceutical interventions based on epidemiological indicators,” *Journal of Mathematical Biology*, vol. 83, p. 42, oct 2021.
- [6] E. Mathieu, H. Ritchie, E. Ortiz-Ospina, M. Roser, J. Hasell, C. Appel, C. Giattino, and L. Rodés-Guirao, “A global database of COVID-19 vaccinations,” *Nature Human Behaviour*, vol. 5, pp. 947–953, jul 2021.
- [7] N. G. Davies, P. Klepac, Y. Liu, K. Prem, M. Jit, C. A. B. Pearson, B. J. Quilty, A. J. Kucharski, H. Gibbs, S. Clifford, A. Gimma, K. van Zandvoort, J. D. Munday, C. Diamond, W. J. Edmunds, R. M. G. J. Houben, J. Hellewell, T. W. Russell, S. Abbott, S. Funk, N. I. Bosse, Y. F. Sun, S. Flasche, A. Rosello, C. I. Jarvis, and R. M. Eggo, “Age-dependent effects in the transmission and control of COVID-19 epidemics,” *Nature Medicine*, vol. 26, pp. 1205–1211, aug 2020.

- [8] R. Verity, L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-Dannenburg, H. Thompson, P. G. T. Walker, H. Fu, A. Dighe, J. T. Griffin, M. Baguelin, S. Bhatia, A. Boonyasiri, A. Cori, Z. Cucunubá, R. FitzJohn, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, D. Laydon, G. Nedjati-Gilani, S. Riley, S. van Elsland, E. Volz, H. Wang, Y. Wang, X. Xi, C. A. Donnelly, A. C. Ghani, and N. M. Ferguson, “Estimates of the severity of coronavirus disease 2019: a model-based analysis,” *The Lancet Infectious Diseases*, vol. 20, pp. 669–677, jun 2020.
- [9] Z. Wang and K. Tang, “Combating COVID-19: health equity matters,” *Nature Medicine*, vol. 26, pp. 458–458, apr 2020.
- [10] F. Ahmed, N. Ahmed, C. Pissarides, and J. Stiglitz, “Why inequality could spread COVID-19,” *The Lancet Public Health*, vol. 5, no. 5, p. e240, 2020.
- [11] S. Lai, N. W. Ruktanonchai, L. Zhou, O. Prosper, W. Luo, J. R. Floyd, A. Wesolowski, M. Santillana, C. Zhang, X. Du, H. Yu, and A. J. Tatem, “Effect of non-pharmaceutical interventions to contain COVID-19 in China,” *Nature*, vol. 585, no. 7825, pp. 410–413, 2020.
- [12] M. U. G. Kraemer, C.-H. Yang, B. Gutierrez, C.-H. Wu, B. Klein, D. M. Pigott, L. du Plessis, N. R. Faria, R. Li, W. P. Hanage, J. S. Brownstein, M. Layan, A. Vespignani, H. Tian, C. Dye, O. G. Pybus, and S. V. Scarpino, “The effect of human mobility and control measures on the COVID-19 epidemic in China,” *Science*, vol. 368, pp. 493–497, may 2020.
- [13] M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. Pastore y Piontti, K. Mu, L. Rossi, K. Sun, C. Viboud, X. Xiong, H. Yu, M. Elizabeth Halloran, I. M. Longini, and A. Vespignani, “The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak,” *Science*, vol. 368, no. 6489, pp. 395–400, 2020.
- [14] P. Nouvellet, S. Bhatia, A. Cori, K. E. C. Ainslie, M. Baguelin, S. Bhatt, A. Boonyasiri, N. F. Brazeau, L. Cattarino, L. V. Cooper, H. Coupland, Z. M. Cucunuba, G. Cuomo-Dannenburg, A. Dighe, B. A. Djaafara, I. Dorigatti, O. D. Eales, S. L. van Elsland, F. F. Nascimento, R. G. FitzJohn, K. A. M. Gaythorpe, L. Geidelberg, W. D. Green, A. Hamlet, K. Hauck, W. Hinsley, N. Imai, B. Jeffrey, E. Knock, D. J. Laydon, J. A. Lees, T. Mangal, T. A. Mellan, G. Nedjati-Gilani, K. V. Parag, M. Pons-Salort, M. Ragonnet-Cronin, S. Riley, H. J. T. Unwin, R. Verity, M. A. C. Vollmer, E. Volz, P. G. T. Walker, C. E. Walters, H. Wang, O. J. Watson, C. Whittaker, L. K. Whittles, X. Xi, N. M. Ferguson, and C. A. Donnelly, “Reduction in mobility and COVID-19 transmission,” *Nature Communications*, vol. 12, p. 1090, dec 2021.
- [15] S. Chang, E. Pierson, P. W. Koh, J. Gerardin, B. Redbird, D. Grusky, and J. Leskovec, “Mobility network models of COVID-19 explain inequities and inform reopening,” *Nature*, vol. 589, no. 7840, pp. 82–87, 2021.
- [16] C. Field, V. Barros, T. Stocker, D. Qin, D. Dokken, K. Ebi, M. Mastrandrea, K. Mach, G. Plattner, and S. Allen, “A special report of working groups I and II of the intergovernmental panel on climate change,” *Managing the risks of extreme events and disasters to advance climate change adaptation*, 2012.

- [17] D. Bichara, Y. Kang, C. Castillo-Chavez, R. Horan, and C. Perrings, *SIS and SIR Epidemic Models Under Virtual Dispersal*, vol. 77. Springer US, 2015.
- [18] D. Bichara and A. Iggidr, “Mathematical Biology Multi-patch and multi-group epidemic models: a new framework,” *Journal of Mathematical Biology*, vol. 77, pp. 107–134, 2018.
- [19] J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G. S. Tomba, J. Wallinga, J. Heijne, M. Sadkowska-Todys, M. Rosinska, and W. J. Edmunds, “Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases,” *PLoS Medicine*, vol. 5, p. e74, mar 2008.
- [20] K. Prem, A. R. Cook, and M. Jit, “Projecting social contact matrices in 152 countries using contact surveys and demographic data,” *PLOS Computational Biology*, vol. 13, p. e1005697, sep 2017.
- [21] K. Prem, Y. Liu, T. W. Russell, A. J. Kucharski, R. M. Eggo, N. Davies, S. Flasche, S. Clifford, C. A. Pearson, J. D. Munday, S. Abbott, H. Gibbs, A. Rosello, B. J. Quilty, T. Jombart, F. Sun, C. Diamond, A. Gimma, K. van Zandvoort, S. Funk, C. I. Jarvis, W. J. Edmunds, N. I. Bosse, J. Hellewell, M. Jit, and P. Klepac, “The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study,” *The Lancet Public Health*, vol. 5, no. 5, pp. e261–e270, 2020.
- [22] R. Kitamura, “An evaluation of activity-based travel analysis,” mar 1988.
- [23] K. W. Axhausen and T. Gärling, “Activity-based approaches to travel analysis: Conceptual frameworks, models, and research problems: Foreign summaries,” *Transport Reviews*, vol. 12, no. 4, pp. 323–341, 1992.
- [24] M. Munizaga, S. Jara-Díaz, J. Olguín, and J. Rivera, “Generating twins to build weekly time use data from multiple single day OD surveys,” *Transportation*, vol. 38, no. 3, pp. 511–524, 2011.
- [25] R. Sameni, “Mathematical Modeling of Epidemic Diseases; A Case Study of the COVID-19 Coronavirus,” mar 2020.
- [26] E. L. Piccolomini and F. Zama, “Monitoring Italian COVID-19 spread by a forced SEIRD model,” *PLoS ONE*, vol. 15, no. 8 August, pp. 1–17, 2020.
- [27] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Journal of Fluids Engineering, Transactions of the ASME*, vol. 82, no. 1, pp. 35–45, 1960.
- [28] F. Auger, M. Hilaret, J. M. Guerrero, E. Monmasson, T. Orłowska-Kowalska, and S. Katsura, “Industrial Applications of the Kalman Filter: A Review,” *IEEE Transactions on Industrial Electronics*, vol. 60, pp. 5458–5471, dec 2013.
- [29] A. Hasan, H. Susanto, V. R. Tjahjono, R. Kusdiantara, E. R. M. Putri, P. Hadisoemarto, and N. Nuraini, “A new estimation method for COVID-19 time-varying reproduction number using active cases,” jun 2020.
- [30] J. Song, H. Xie, B. Gao, Y. Zhong, C. Gu, and K.-S. Choi, “Maximum likelihood-based extended Kalman filter for COVID-19 prediction,” *Chaos, Solitons & Fractals*, vol. 146, p. 110922, 2021.

- [31] M. Olivares, M. Goic, G. Weintraub, J. Covarrubia, and C.án@, “El dispar impacto de las cuarentenas en las comunas de la Región Metropolitana,” pp. 1–5, 2020.
- [32] G. E. Mena, P. P. Martinez, A. S. Mahmud, P. A. Marquet, C. O. Buckee, and M. Santillana, “Socioeconomic status determines COVID-19 incidence and related mortality in Santiago, Chile,” *Science*, vol. 5298, no. April, p. eabg5298, 2021.
- [33] M. Bennett, “All things equal? Heterogeneity in policy effectiveness against COVID-19 spread in chile,” *World Development*, vol. 137, p. 105208, 2021.
- [34] N. Gozzi, M. Tizzoni, M. Chinazzi, L. Ferres, A. Vespignani, and N. Perra, “Estimating the effect of social inequalities on the mitigation of COVID-19 across communities in Santiago de Chile,” *Nature Communications*, vol. 12, no. 1, pp. 1–9, 2021.
- [35] MINCIENCIA and MINSAL, “MinCiencia/Datos-COVID19.”
- [36] Z. Cai and D. Zhao, “Unscented Kalman filter for non-linear estimation,” *Geomatics and Information Science of Wuhan University*, vol. 31, no. 2, pp. 180–183, 2006.
- [37] M. Katzfuss, J. R. Stroud, and C. K. Wikle, “Understanding the Ensemble Kalman Filter,” *American Statistician*, vol. 70, no. 4, pp. 350–357, 2016.
- [38] Y. Shikhmurzaev and V. Shikhmurzaev, “A unified activities-based approach to the modelling of viral epidemics and COVID-19 as an illustrative example,” *medRxiv*.
- [39] D. J. Daley and J. Gani, *Epidemic Modelling*. Cambridge University Press, feb 1984.
- [40] F. Brauer, C. Castillo-Chavez, and Z. Feng, *Mathematical Models in Epidemiology*, vol. 69 of *Texts in Applied Mathematics*. New York, NY: Springer New York, 2019.
- [41] C. Cosner, J. C. Beier, R. S. Cantrell, D. Impoinvil, L. Kapitanski, M. D. Potts, A. Troyo, and S. Ruan, “The effects of human movement on the persistence of vector-borne diseases,” *Journal of Theoretical Biology*, vol. 258, no. 4, pp. 550–560, 2009.
- [42] Y.-H. Hsieh, P. Van Den Driessche, and L. Wang, “Impact of Travel Between Patches for Spatial Spread of Disease,” *Bulletin of Mathematical Biology*, vol. 69, pp. 1355–1375, 2007.
- [43] S. Ruan, W. Wang, and S. A. Levin, “The effect of global travel on the spread of SARS,” *Mathematical Biosciences and Engineering*, vol. 3, no. 1, pp. 205–218, 2006.
- [44] R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control*. OUP Oxford, 1992.
- [45] J. Wallinga, P. Teunis, and M. Kretzschmar, “Using Data on Social Contacts to Estimate Age-specific Transmission Parameters for Respiratory-spread Infectious Agents,” *American Journal of Epidemiology*, vol. 164, pp. 936–944, nov 2006.
- [46] W. Edmunds, G. Kafatos, J. Wallinga, and J. Mossong, “Mixing patterns and the spread of close-contact infectious diseases,” *Emerging Themes in Epidemiology*, vol. 3, p. 10, dec 2006.

- [47] G. Y. Kulikov and M. V. Kulikova, “Accurate Numerical Implementation of the Continuous-Discrete Extended Kalman Filter,” *IEEE Transactions on Automatic Control*, vol. 59, pp. 273–279, jan 2014.
- [48] B. Anderson and J. Moore, *Optimal Filtering*. Dover Publications, 2005.
- [49] D. Simon, *Optimal State Estimation*. Hoboken, NJ, USA: John Wiley & Sons, Inc., may 2006.
- [50] H. E. Rauch, F. Tung, and C. T. Striebel, “Maximum likelihood estimates of linear dynamic systems,” *AIAA Journal*, vol. 3, pp. 1445–1450, aug 1965.
- [51] D. Simon, “Kalman filtering with state constraints: a survey of linear and nonlinear algorithms,” *IET Control Theory & Applications*, vol. 4, pp. 1303–1318, aug 2010.
- [52] R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, and J. Shaman, “Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2),” *Science*, vol. 368, pp. 489–493, may 2020.
- [53] O. Byambasuren, M. Cardona, K. Bell, J. Clark, M.-L. McLaws, and P. Glasziou, “Estimating the extent of asymptomatic COVID-19 and its potential for community transmission: Systematic review and meta-analysis,” *Official Journal of the Association of Medical Microbiology and Infectious Disease Canada*, vol. 5, pp. 223–234, dec 2020.
- [54] Z. Gao, Y. Xu, C. Sun, X. Wang, Y. Guo, S. Qiu, and K. Ma, “A systematic review of asymptomatic infections with COVID-19,” *Journal of Microbiology, Immunology and Infection*, vol. 54, pp. 12–16, feb 2021.
- [55] S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, and J. Lessler, “The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application,” *Annals of Internal Medicine*, vol. 172, pp. 577–582, may 2020.
- [56] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T. T. Lam, J. T. Wu, G. F. Gao, B. J. Cowling, B. Yang, G. M. Leung, and Z. Feng, “Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia,” *New England Journal of Medicine*, vol. 382, pp. 1199–1207, mar 2020.
- [57] A. Tariq, E. A. Undurraga, C. C. Laborde, K. Vogt-Geisse, R. Luo, R. Rothenberg, and G. Chowell, “Transmission dynamics and control of COVID-19 in Chile, March-October, 2020,” *PLOS Neglected Tropical Diseases*, vol. 15, p. e0009070, jan 2021.
- [58] MINSAL, “Se inicia proceso de vacunación masiva contra COVID-19 - Ministerio de Salud - Gobierno de Chile,” 2021.
- [59] P. d. v. y. transporte urbano SECTRA, “Encuesta origen y destino de viajes 2012,” p. 84, 2014.

- [60] TransporTeinforma, “MTT utilizará Big Data para recopilar información y mejorar las políticas de movilidad.”
- [61] A. Carranza, M. Goic, E. Lara, M. Olivares, G. Y. Weintraub, J. Covarrubia, C. Escobedo, N. Jara, and L. J. Basso, “The Social Divide of Social Distancing: Shelter-in-Place Behavior in Santiago during the Covid-19 Pandemic,” sep 2020.
- [62] Y. Xiang, Y. Jia, L. Chen, L. Guo, B. Shu, and E. Long, “COVID-19 epidemic prediction and the impact of public health interventions: A review of COVID-19 epidemic models,” *Infectious Disease Modelling*, vol. 6, no. 24, pp. 324–342, 2021.
- [63] D. Kai, G.-P. Goldstein, A. Morgunov, V. Nangalia, and A. Rotkirch, “Universal Masking is Urgent in the COVID-19 Pandemic: SEIR and Agent Based Models, Empirical Validation, Policy Recommendations,” 2020.
- [64] S. Jara-Díaz, M. Munizaga, and J. Olguín, “The role of gender, age and location in the values of work behind time use patterns in Santiago, Chile,” *Papers in Regional Science*, vol. 92, no. 1, pp. 87–102, 2013.
- [65] N. M. Ferguson, D. Laydon, G. Nedjati-Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhattia, A. Boonyasiri, Z. Cucunubá, G. Cuomo-Dannenburg, A. Dighe, I. Dorigatti, H. Fu, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, L. C. Okell, S. Van Elsland, H. Thompson, R. Verity, E. Volz, H. Wang, Y. Wang, P. Gt Walker, C. Walters, P. Winskill, C. Whittaker, C. A. Donnelly, S. Riley, and A. C. Ghani, “Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand,” 2020.
- [66] G. Y. Kulikov and M. V. Kulikova, “The continuous–discrete extended Kalman filter revisited,” *Russian Journal of Numerical Analysis and Mathematical Modelling*, vol. 32, pp. 27–38, jan 2017.
- [67] B. Cazelles and N. Chau, “Using the Kalman filter and dynamic models to assess the changing HIV/AIDS epidemic,” *Mathematical Biosciences*, vol. 140, pp. 131–154, mar 1997.
- [68] D. Ndanguza, I. S. Mbalawata, H. Haario, and J. M. Tchuente, “Analysis of bias in an Ebola epidemic model by extended Kalman filter approach,” *Mathematics and Computers in Simulation*, vol. 142, pp. 113–129, 2017.
- [69] A. Torres-Signes and J. A. Dip, “A Bayesian Functional Methodology for Dengue Risk Mapping in Latin America and the Caribbean,” *Acta Tropica*, vol. 215, no. April 2020, p. 105788, 2021.
- [70] P. Narula, V. Piratla, A. Bansal, S. Azad, and P. Lio, “Parameter estimation of tuberculosis transmission model using Ensemble Kalman filter across Indian states and union territories,” *Infection, Disease and Health*, vol. 21, no. 4, pp. 184–191, 2016.
- [71] Q. Yang, C. Yi, A. Vajdi, L. W. Cohnstaedt, H. Wu, X. Guo, and C. M. Scoglio, “Short-term forecasts and long-term mitigation evaluations for the COVID-19 epidemic in Hubei Province, China,” *Infectious Disease Modelling*, vol. 5, pp. 563–574, 2020.

- [72] C. H. Nkwayep, S. Bowong, J. J. Tewa, and J. Kurths, “Short-term forecasts of the COVID-19 pandemic: a study case of Cameroon,” *Chaos, Solitons and Fractals*, vol. 140, p. 110106, 2020.
- [73] F. Arroyo-Marioli, F. Bullano, S. Kucinskis, and C. Rondón-Moreno, “Tracking R of COVID-19: A new real-time estimation using the Kalman filter,” *PLoS ONE*, vol. 16, no. 1 January, pp. 1–51, 2021.
- [74] A. Gomez-Exposito, J. A. Rosendo-Macias, and M. A. Gonzalez-Cagigal, “Monitoring and Tracking the Evolution of a Viral Epidemic Through Nonlinear Kalman Filtering: Application to the Covid-19 Case,” *IEEE Journal of Biomedical and Health Informatics*, pp. 3–14, 2021.
- [75] A. Rajaei, M. Raeiszadeh, V. Azimi, and M. Sharifi, “State estimation-based control of COVID-19 epidemic before and after vaccine development,” *Journal of Process Control*, vol. 102, pp. 1–14, 2021.
- [76] R. Bansal, A. Kumar, A. K. Singh, and S. Kumar, “Stochastic filtering based transmissibility estimation of novel coronavirus,” *Digital Signal Processing: A Review Journal*, vol. 112, p. 103001, 2021.
- [77] A. Hasan, E. R. Putri, H. Susanto, and N. Nuraini, “Data-driven modeling and forecasting of COVID-19 outbreak for public policy making,” *ISA Transactions*, no. xxxx, pp. 7–15, 2021.
- [78] A. Hasan and Y. Nasution, “A compartmental epidemic model incorporating probable cases to model COVID-19 outbreak in regions with limited testing capacity,” *ISA Transactions*, no. xxxx, 2021.
- [79] C. Rackauckas and Q. Nie, “DifferentialEquations.jl – A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia,” *Journal of Open Research Software*, vol. 5, p. 15, may 2017.
- [80] Y. Ma, S. Gowda, R. Anantharaman, C. Laughman, V. Shah, and C. Rackauckas, “ModelingToolkit: A Composable Graph Transformation System For Equation-Based Modeling,” mar 2021.
- [81] La Tercera, “[PRENSA] Estudio muestra la distribución de adultos mayores en Santiago – COES.”
- [82] Y. Ma, S. Gowda, R. Anantharaman, C. Laughman, V. Shah, and C. Rackauckas, “ModelingToolkit: A Composable Graph Transformation System For Equation-Based Modeling,” 2021.

Anexo: Clasificación IPS 2019

Categoría	Rk	Comuna	IPS 2019
ALTA PRIORIDAD SOCIAL	1	La Pintana	83.03
	2	Lo Espejo	81.78
	3	Cerro Navia	81.04
	4	San Ramón	80.28
	5	Isla de Maipo	80.28
	6	María Pinto	78.24
MEDIA ALTA PRIORIDAD SOCIAL	7	Curacaví	77.39
	8	Conchalí	77.06
	9	El Monte	76.85
	10	Paine	75.91
	11	Melipilla	75.80
	12	Buín	75.77
	13	Lo Prado	75.54
	14	Padre Hurtado	75.37
	15	San Joaquín	75.16
	16	San Bernardo	75.14
	17	El Bosque	74.63
	18	San José de Maipo	74.52
	19	Recoleta	73.84
MEDIA BAJA PRIORIDAD SOCIAL	20	Independencia	71.36
	21	Til-Til	71.21
	22	Alhué	70.90
	23	Lampa	70.85
	24	Quinta Normal	70.23
	25	La Granja	70.14
	26	Estación Central	69.89
	27	Pedro Aguirre Cerda	69.49
	28	Peñaflor	68.99
	29	San Pedro	68.75
	30	Renca	68.53
	31	Talagante	68.45
	32	La Cisterna	66.93
	33	Pirque	66.12
	34	Pudahuel	65.05

Categoría	Rk	Comuna	IPS 2019
BAJA PRIORIDAD SOCIAL	35	Colina	64.37
	36	Cerrillos	64.36
	37	Calera de Tango	64.34
	38	Puente Alto	64.27
	39	Peñalolén	64.12
	40	La Florida	59.28
	41	Maipú	58.39
	42	Huechuraba	58.06
	43	Santiago	57.11
	44	Quilicura	56.01
SIN PRIORIDAD SOCIAL	45	Macul	55.02
	46	San Miguel	53.34
	47	Nuñoa	37.36
	48	La Reina	32.48
	49	Lo Barnechea	32.00
	50	Providencia	26.63
	51	Las Condes	11.68
	52	Vitacura	6.26

Tabla 1: Clasificación de las comunas de la Región Metropolitana de Santiago de acuerdo al Índice de Prioridad Social (IPS) 2019. Fuente: Seremi de Desarrollo Social y Familia Metropolitana [1].