Universidad de Concepción
Dirección de Postgrado
Facultad de Ciencias Físicas y Matemáticas
Programa de Doctorado en Ciencias Aplicadas
con Mención en Ingeniería Matemática

IDENTIFICACIÓN DE PARÁMETROS Y APLICACIONES DE MODELOS FENOMENOLÓGICOS Y MECANÍSTICOS DE BROTES EPIDÉMICOS

(PARAMETER IDENTIFICATION AND APPLICATIONS OF PHENOMENOLOGICAL AND MECHANISTIC MODELS OF EPIDEMIC OUTBREAKS)

Tesis para optar al grado de
Doctor en Ciencias Aplicadas con mención en Ingeniería Matemática

Leidy Yissedt Lara Díaz

CONCEPCIÓN-CHILE
2022

Profesor Guía: Dr. Raimund Bürger
CI^{2}MA y Departamento de Ingeniería Matemática
Universidad de Concepción, Chile

Cotutor: Dr. Gerardo Chowell
School of Public Health
Georgia State University

Parameter identification and application of phenomenological and mechanistic models of epidemic outbreaks

Leidy Yissedt Lara Díaz

Directores de Tesis: Dr. Raimund Bürger, Universidad de Concepción, Chile.
Dr. Gerardo Chowell, Georgia State University.

Director de Programa: Dr. Raimund Bürger, Universidad de Concepción, Chile.

Comisión evaluadora

Prof. .

Prof.

Prof.

Prof.

Comisión examinadora

Firma:
Prof.

Firma:
Prof.

Firma:
Prof.

Firma:
Prof.

Calificación:

Concepción, xx de xx de 2022

Abstract

This thesis is focused on models given by ordinary differential equations (ODEs) to describe the temporal dynamics of epidemic outbreaks. Such studies and applications involve the use of databases, statistical tools, and numerical simulations. The current pandemic situation of the spread of the SARS-Cov2 virus has motivated us to dedicate part of this work to the modeling of COVID-19 in Chile and Colombia. Specifically, the work considers two types of epidemiologic models to capture the dynamics of growth and spread of infectious diseases. Within these two types, we have phenomenological models and mechanistic models. Specifically to the first type of models we develop a comparative analysis and an analysis of the growth curves obtained for the case of COVID-19 in Colombia. On the other hand, the second type of models is utilized to describe initial outbreak COVID-19 in Chile.

As a quick overview of the thesis, we began by studying four different phenomenological models, which we compared using 37 databases and different statistical measures to define which of these models better capture the different dynamics, such development is exposed in **Chapter 1**. In **Chapter 2**, to complement this comparative analysis, we extended the study using synthetic databases and statistical tools to see which model is closer to the others, trying to capture the advantages or disadvantages of each when fitting growth curves. **Chapter 3** is developed with COVID-19 data from Colombia. Various phenomenological models are applied to capture the growth curves to national and regional levels, and then some short-term forecasts are obtained. Besides, from the Colombia epidemic curves generated by GGM, the effective reproduction number R_t is computed. All this is to understand the transmission dynamics of COVID-19 in Colombia in both the early and final phases. In **Chapter 4**, we propose a compartmental mechanistic model inspired by the initial spread of COVID-19 in Chile, to incorporate an identifiability study, using synthethic and regional Chilean data. In **Chapter 5** we summarize the main conclusions obtained in the previous chapters and the topics that could be addressed in future works. Finally, in **Appendices A** and **B**, we include Matlab programs, figures, and tables that complement different results from our work.

In summary, the general objective of the thesis is divided into the following four specific objectives that define the content of each of the developed chapters;

The first objective of this thesis is to compare the most common phenomenological models to capture epidemic growths and see which are the most versatile to capture these outbreaks and

see if the ability to fit well depends on the number of parameters that define each model.

The second objective of this thesis is, having already established a comparison between different phenomenological models for the adjustment of real data, to quantify the distance between two models as a measure of differences in the dynamics that each model is capable of generating.

The third objective involves the application of phenomenological models to the study of the epidemic growth dynamics presented in Colombia for the COVID-19 disease to analyze the growth rates and the short-term forecasts, also involving a study to compare the effective reproduction numbers obtained both at the national and regional levels.

The fourth and final objective is to propose an identifiability analysis for the parameters of a mechanistic model by compartments that study the propagation dynamics in the Chilean regions for the SARS-CoV2 virus, considering the initial period of the outbreak and the quarantine measures independently applied in each Chilean region.

Resumen

Esta tesis se centra en modelos definidos por ecuaciones diferenciales ordinarias (EDOs) para describir la dinámica temporal de los brotes epidémicos, los cuales son estudiados y aplicados para comparar y modelar los brotes epidémicos de nuestro interés. Tales estudios y aplicaciones involucran el uso de bases de datos, herramientas estadísticas y simulaciones o implementaciones numéricas. La actual situación de pandemia por la propagación del virus SARS-Cov2, nos permitimos dedicar parte de este trabajo a la modelización de COVID-19 en Chile y Colombia. Específicamente, el trabajo considera dos tipos de modelos epidemiológicos para ajustar o capturar la dinámica de crecimiento y propagación de enfermedades infecciosas, dentro de estos dos tipos, contamos con modelos fenomenológicos y modelos mecanicistas, donde específicamente a los primeros modelos se desarrolla un análisis comparativo y un análisis de las curvas de crecimiento obtenidas para el caso de COVID-19 en Colombia, por otro lado, el segundo tipo de modelos, lo aplicamos estudiar la identificabilidad de los parámetros del modelo para el brote inicial de COVID-19 en Chile.

Haciendo un vista general de la tesis, comenzamos estudiando cuatro diferentes modelos fenomenológicos, los cuales comparamos usando 37 bases de datos reales y diferentes medidas estadísticas para definir cuál de estos modelos logra capturar mejor diferentes dinámicas, tal desarrollo es expuesto en el **Capítulo 1**. En el **Capítulo 2**, para ampliar este análsis comparativo, extendimos el estudio usando bases de datos sintéticos, y herramientas estadísticas para ver qué modelo se aproxima más a los otros, tratando de capturar las ventajas o desventajas de cada uno al momento de ajustar curvas de crecimiento. El **Capítulo 3** se desarrolla con datos de COVID-19 de Colombia, donde se emplean diferentes modelos fenomenológicos para capturar las dinámicas de crecimiento a nivel nacional y regional, que permiten hacer proyecciones a corto plazo. Además de las curvas epidémicas generadas para Colombia por el método de crecimiento generalizado (GGM), el número efectivo de reproducción es calculado. Todo esto para entender la dinámica de transmisión del COVID-19 en Colombia tanto en la fase inicial como el la final considerada en el estudio. En el **Capítulo 4** proponemos un modelo mecanístico compartimental inspirado en la propagación inicial del COVID-19 en Chile, con el cual buscamos incorporar un estudio de identificabilidad usando datos sintéticos y datos regionales chilenos. En el **Capítulo 5** resumimos las principales conclusiones obtenidas en cada unos los capítulos anteriores, así como los temas que pueden ser extendidos y los posibles trabajos futuros. Finalmente en los **Apéndices A** y **B**, incluimos algunos programas de MATLAB, figuras y tablas

que complementan differentes resultados de nuestro trabajo.

En resumen el objetivo general de la tesis se divide en los siguientes cuatro objetivos específicos que definen el contenido de cada uno de los capítulos desarrollados;

El primer objetivo de esta tesis es comparar los modelos fenomenológicos más comunes para capturar los crecimientos epidémicos y ver cuáles son los más versátiles para capturar los brotes epidémicos, y ver si la capacidad de buen juste depende de la cantidad de parámetros que definen cada modelo.

El segundo objetivo de esta tesis es, ya teniendo establecida una comparación entre diferentes modelos fenomenológicos para el ajuste de datos reales, se busca cuantificar que tan distantes son estos modelos entre sí, midiendo las diferencias en las dinámicas que cada modelo fenomenológico es capaz de generar.

El tercer objetivo consiste en la aplicación de modelos fenomenológicos al estudio de la dinámica de crecimiento epidémico presentado en Colombia por la enfermedad COVID-19 para analizar las tasas de crecimiento y los pronósticos a corto plazo, involucrando además un estudio para comparar los números de reproducción efectivos obtenidos tanto a nivel nacional como regional.

El cuarto y último objetivo radica en proponer estudio de identificabilidad a los parámetros de un modelo mecanístico por compartimentos para estudiar las dinámicas de propagación en las regiones de Chile para el virus SARS-CoV2, teniendo en cuenta el periodo inicial del brote y las medidas de cuarentena aplicadas en cada región, las cuales se dieron de manera independiente.

Agradecimientos

Estos últimos 2 años del doctorado han estado marcados por muchos movimientos que han forzado muchos cambios y nuevas adaptaciones, los recientes estallidos sociales, tanto en Chile como en Colombia, que traen a la memoria muchos de los pendientes o deudas a la sociedad y al medio ambiente, por otro lado, la pandemia provocada por el virus de SARS-Cov2, que ha puesto a prueba a toda la humanidad, y que nos hace replantear nuestra posición y nuestra función en el planeta. Cambios, resistencias, miedos, pérdidas, y nuevas oportunidades es lo que ha traído estos últimos tiempos a mi vida, y han hecho que mi trabajo tenga un motivo y una mayor razón para aportar. Así que este trabajo no podía ser posible sin las personas que me soportaron y apoyaron en todos estos años de avance y cambios. Es por ello que quiero agredecer a mis padres Edith y Juan, quienes son mi pilar y me han brindado todo su amor y ejemplo, para enfrentar la vida, para creer en mí, y para perseguir mis sueños, a mis hermanos Eimmy y Juan, por ser mis cómplices y soporte en momentos de duda, a mi sobrino Mattias, quien siempre me regala momentos de juego y risas, a mi amado esposo Gabriel, por motivar mi proyecto de vida, y por ser mi luz y razón de sonreir cada día.

De manera especial quiero igualmente agradecer a mi director de tesis, el profesor Raimund Bürger, a quien admiro por su disciplina, buena volutad y dispocisión. Gracias por aceptar ser mi guía y por apoyar cada una de mis etapas en el doctorado. Cada uno de sus consejos me permitieron sobrellevar momentos de dificultad y poder ver el potencial en cada tema que trabajamos.

A mi co-tutor, Gerardo Chowell por su amabilidad y por ser mi guía en este mundo de la epidemiología matemática, que me permitieron comprender varios de los problemas y mecanísmos que pueden ser útiles para solucionarlos, más aún de priorizar los enfoques en la actual situación de pandemia que nos desafía día a día. También quiero agradecerle por el buen recibimiento y acogida que me brindó junto a su familia durante mi estancia de pasantía.

A mis colaboradores, el investigador Ilja Kröker, por aceptar trabajar con nosotros y brindarnos su apoyo en el desarrollo de la implementación del modelo para el estudio del COVID-19 y sus aplicaciones con datos de Chile, y a la investigadora Amna Tariq, quien nos invitó a participar en un estudio para la modelación del COVID-19 en Colombia.

Al Centro de Investigación en Ingeniería Matemática (CI^{2}MA) de la Universidad de Concepción, por brindarme el espacio y las instalaciones para poder trabajar comodamente. Al personal administrativo, principalmente a la Sra. Lorena por brindar siempre un ambiente amigable y casero en los espacios compartidos. Al departamento de Ingeniería Matemática, en especial a la secretaria Cecilia Leiva y al técnico José Parra, por sus apoyos.

A los profesores del programa especialmente a los profesores, Manuel Solano y Rodolfo Rodriguez, quienes han estado siempre dispuestos a brindar su ayuda en momentos de dificultad, a mis compañeros del programa de doctorado por todas esas conversaciones con cafés y postre.

En especial a Cynthia, Néstor, Adrian y Yolanda por la amistad y compañía en la fase inicial del doctorado, a Daniel, Cristian, Joaquín, Bryan, Paul, Rafael, Romel, Juan Pablo y Wiiliam, por cada consejo y ayuda.

Agradecer además a todas las instituciones que me permitieron realizar este trabajo: a la Universidad de Concepción con la Beca de estipendio durante el primer año, a la Agencia Nacional de Investivación y Desarrollo (ANID) a través de su programa de becas ANID-PCHA/Doctorado Nacional/201921190640 y a todos los beneficios otorgados. Así como a los proyectos ANID/PIA/ACT210030 y FONDECYT 1170473 - 1210610.

Finalmente, y no menos importante a mis abuelos, tías, tíos, cuñada y primos, quienes han estado presentes en cada momento de mi vida, gracias por hacerme sentir importante, así como a mis suegros y cuñados, quienes me han hecho sentir en familia desde que llegué a Chile.

Leidy Yissedt Lara Díaz

Contents

List of Tables

List of Figures

Introduction

The COVID-19 pandemic has become one of thee main topics of daily news around the world during the last few years. We recall that the coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was declared a global pandemic by the World Health Organization (WHO) on March 11, 2020 [28, 185]. This highly contagious virus has impacted governments, public institutions, and stressed health care systems, confining people to their homes and causing country-wide lockdowns resulting in a global economic crisis. This situation has made evident the great importance of theories, methodologies, techniques, and new models that allow progress in the creation of theoretical and quantitative frameworks that guide scientists and health control agencies to understand the epidemiological dynamics present in a population. Moreover these frameworkers allows to explore possible scenarios by applying some control measures, such as the use of face masks [31, 108], quarantines [26, 81], and vaccination campaigns [20, 21, 49].

Various models are used for such studies, but we focus on phenomenological and mechanistic models, both described mainly by ordinary differential equations (ODEs). The phenomenological growth models (PGMs) help capture the dynamics of epidemic growth in a simple framework. Their description involves few parameters, which often allow an explicit solution to be obtained. Their equations have an empirical approach, i.e., without an explicit basis of physical laws or mechanisms, which are often difficult to identify, making them a very efficient and fast tool for forecasting with identifiable parameters. On the other hand, mechanistic models attempt to describe the transmission of the disease in a population represented by the infection states by compartments. Such an idea and characterization is developed thanks to the well-known work of Kermack and McKendrick [85]. Within the compartmental models, [2, 13, 61, 96, 176, 188] the total population is subdivided into at least two compartments or epidemiological states (which can be susceptible and infected; but many others can be considered). It is necessary to specify the rates of progression between compartments, as well as the incidence and possibly of births and deaths of individuals, that is, in this case, biological mechanisms are involved, being the systems of ordinary differentiable equations (ODEs) those that describe the progression of the epidemic. These models are often described with several parameters that define the dynamics and biological phenomena. It is difficult to obtain an explicit solution, but they have been very useful for estimating parameters of interest with the help of real data and the exploration of possible scenarios.

Together with the models described previously, it is possible to define a parameter of great interest in mathematical epidemiology, the basic reproductive number R_0. Epidemiologically R_0 gives the number of secondary infectious cases generated by a fully susceptible primary infectious individual during the early transmission when the population is in the absence of control interventions. This parameter plays a role of a threshold value for the dynamics of the system, with which is possible to determine in a population whether or not there will be an epidemic. Therefore these type of studies are of great interest and importance because all the information that can be inferred with the help of models and data, being low-cost, fast and timely tools when making decisions.

Next, we introduce the three big questions that motivate our work and then we will give a description to solve them,

1) Growth dynamics models provide an important quantitative framework for characterizing epidemic trajectories, generating estimates of key parameters, evaluating the impact of control interventions, obtaining information on the contribution of different transmission routes, and producing short- and long-term forecasts. These advantages motivated the following question: can the most appropriate growth model be chosen for a given epidemic? Which focused our first purpose in **Chapter 1** and **2**, trying to shed light on the performance of different growth models in describing real epidemic outbreaks. Specifically, in **Chapter 1**, we employ four different growth models based on differential equations (two with two parameters and two with three parameters), and we examine them using 37 databases of different infectious outbreaks. That consists of a time series of incidence cases to identify the best model to describe epidemic growth in each case. On the other hand, **Chapter 2** attempts to answer which epidemic growth model is better for capturing the dynamics generated by other growth models? For this, a methodology is created that helps quantify the differences between the dynamics obtained from different models that capture processes of epidemic growth.

2) With the emergency presented by COVID-19 and the knowledge gathered with the phenomenological growth models, some questions appear, such as, for the COVID-19 disease, do the generalized models capture better than simple models? Does the new sub-epidemic model capture the multiple peaks evidenced by this epidemic? How good will their short-term forecast be? What other contributions can the fitting curves from phenomenological growth models make? This concern led us to work together with other colleagues. Using the Colombia epidemic data, we apply various phenomenological models to compare their fits and forecast and employ performance metrics between the data and the models to determine the quality of fits. The effective reproduction number also is computed to understand the epidemic impact evidenced in Colombia. Such development is presented in **Chapter 3**.

3) Knowing the compartmental epidemiological models, and the situation experimented in Chile in the initial period, where quarantines were applied in different territorial units,

with different purposes, we are wondering what model of this type can help model the situation presented in Chile, that involves the implemental quarantines, and allows us to measure their impact? A homogeneously mixed compartmental model is proposed with the quarantine dynamics for the Chilean case (combining ideas from [31,81]). We also carry out a identifiability study for some parameter sets to fitting, which we wish to capture for the different and particular strategies applied in each Chilean region. Such advances are showed in **Chapter 4**.

Organization of this thesis

The present thesis is organized as follows:

In **Chapter 1**, for different phenomenological growth models, we propose a comparative analysis between four parametric ODE-based models, namely the logistic and Gompertz model with their respective generalizations that in each case consists in elevating the cumulative incidence function to a power $p \in [0, 1]$. This parameter within the generalized models provides a criterion on the early growth behavior of the epidemic between constant incidence for $p = 0$, sub-exponential growth for $0 < p < 1$ and exponential growth for $p = 1$.

The contents of **Chapter 1** correspond to the article [22]:

- R. Bürger, G. Chowell, and L. Y. Lara-Díaz. (2019). Comparative analysis of phenomenological growth models applied to epidemic outbreaks, *Mathematical Biosciences and Engineering*: MBE, 16(5), 4250–4273. `https://doi.org/10.3934/mbe.2019212`.

In **Chapter 2**, we contribute to a systematic study of differences between models and how such differences may explain the ability of centain models to provide a better fit to data than others . To this end, measures of the distance are defined that describe the differences in the dynamics between different dynamics models. The distance of one growth model from another one quantifies how well the former fits data generated by the latter. This concep of distance is, however, not symmetric. The procedure of calculating distances is applied to synthetic data and real data from influenza, Ebola and COVID-19 outbreaks.

The contents of **Chapter 2** correspond to the article [16]:

- R. Bürger, G. Chowell, and L. Y. Lara-Díaz. (2021). Measuring differences between phenomenological growth models applied to epidemiology, *Mathematical Biosciences*, 334, 108558. `https://doi.org/10.1016/j.mbs.2021.108558`.

In **Chapter 3**, We employ different phenomenological growth models to characterize the COVID-19 outbreak in Colombia. The fits are applied to a national and regional level. Several estimations and fitting curves are obtained, with which the effective reproduction number R_t and sort-term forecasts can be calculated. This work is part of a collaboration with different colleagues, where other concepts and calculations are also included.

The contents of **Chapter 3** correspond to the article [150]:

- A. Tariq, T. Chakhaia, S. Dahal, A. Ewing, X. Hua, S. K. Ofori, O. Prince, A. Salindri, A. E. Adeniyi, J. Banda, P. Skums, R. Luo, **L.Y. Lara-Díaz**, **R. Bürger**, I. C-H. Fung, E. Shim, A. Kirpich, A. Srivastava, **G. Chowell**. (2022). An investigation of spatial-temporal patterns and predictions of the coronavirus 2019 pandemic in Colombia, 2020–2021, *PLOS Neglected Tropical Diseases*, 16(3), e0010228. https://doi.org/10.1371/journal.pntd.0010228.

In **Chapter 4**, we introduce an adaptation to a computational approach to assessing parameter identifiability in compartment models; for this work, we construct a model of the propagation and control of COVID-19 inspired by the situation experimented in Chile at the beginning of 2020. We include an experimental approach to verify if the estimated parameter set to fit the model to a data curve is identifiable. Then, with this idea, we construct a methodology where synthetic data generated from the same model are used to analyze a parameter set of interest, how many parameters and combinations the compartmental model is capable of recovery, and the fit times. With this strategy of mixing the parameter number to estimate and change the fit times, we developed a methodology to determine the cases for which structural and practical identifiability can be guaranteed. Finally, we test the result obtained using synthetic data fitting Chilean data.

The contents of **Chapter 4** correspond to research:

- R. Bürger, G. Chowell, I. Kröker and L. Y. Lara-Díaz, Sensitivity and identifiability analysis for a model of the propagation and control of COVID-19 in Chile, *(in preparation)*.

Introducción

La pandemia de COVID-19 es actualmente el tema principal de las noticias diarias en todo el mundo. Recordemos que la enfermedad por coronavirus 2019 (COVID-19), causada por el síndrome respiratorio agudo severo coronavirus 2 (SARS-CoV-2), fue declarada pandemia mundial por la Organización Mundial de la Salud (OMS) el 11 de marzo de 2020 [28, 185]. Este virus sin precedentes altamente contagioso ha impactado a gobiernos, instituciones públicas y ha estresado los sistemas de atención médica, confinando a las personas en sus hogares y provocando cierres en todos los países, lo que ha generado una crisis económica mundial. Hecho que ha dejamos más en evidencia la gran importancia y responsabilidad que trae el desarrollo de teorías, metodologías, técnicas y nuevos modelos que permitan avanzar en la creación de marcos teóricos y cuantitativos que guíen a los científicos y a los organismos de control sanitarios a entender las dinámicas epideomiológicas presentes en una población, así como el poder explorar posibles escenarios al aplicar ciertass medidas de control, como lo han sido el uso de mascarillas [31, 108], las cuarentenas [26, 81], y las jornadas de vacunación [20, 21, 49].

Muchos modelos han sido usados para tales estudios, pero nosotros nos centramos en los modelos fenomenológicos de crecimiento y los modelos mecanísticos, ambos descritos mayormente con ecuaciones diferenciales ordinarias (EDOs). Los fenomenológicos describen principalmente las dinámicas de crecimiento epidémico ofreciendo una forma simple para su descripción donde involucran pocos parámetros, que muchas veces permiten obtener una solución explícita, y además sus ecuaciones evitan incorporar mecanísmos biológicos, que muchas veces son díficiles de identificar, lo que los hace ser una herramienta muy eficiente y rápida para realizar pronósticos con parámetros identificables. Por su parte, los modelos mecanísticos intentan describir la transmisión de la enfermedad en una población representando los estados de infeccion por compartimentos, tal idea y caracterización se desarrolla gracias al bien conocido trabajo de Kermack y McKendrick [85]. En estos modelos compartimentales [2, 13, 61, 96, 176, 188] la población total es subdividida en al menos dos compartimentos o estados epidemiológicos (que puede ser susceptibles e infectados, pero muchos otros pueden ser considerados), la tasas de progresión entre los compartimentos así como la incidencia y la posibilidad de nacimientos y muertes de individuos, necesitan ser específicados, es decir, en este caso se involucran mecanísmos biológicos, y por tal hecho se trabajan con sistemas de ecuaciones diferenciables ordinarias (EDOs), que muchas veces vienen descritas con un número de parámetros que definen las dinámicas y los fenómenos biológicos, siendo difícil obtener una solución explícita, pero han sido muy útiles para estimar parámetros de interés con ayuda de datos reales, así como la exploración de posibles escenarios.

Con la aplicación de los modelos descritos anteriormente es posible definir un parámetro de mucho interés en la epidemiología matemática, que es el número reproductivo básico R_0, el cual representa el número de casos infecciosos secundarios generados por un individuo infeccioso primario en una población totalmente susceptible, es decir, en ausencia de intervenciones de control, con el cual es posible determinar si habrá o no una epidemia en una poblacion. Es por ello que, este tipo se estudios son de gran interés e importancia, ya que con toda la información que se puede inferir con ayuda de los modelos y los datos, los hace ser una herramienta accesible, de bajo costo, además rápida al momento de tomar de decisiones en momentos de una emegencia.

Introducimos los tres grandes problemas que motivaron nuestro trabajo y luego daremos una descripción para resolverlos,

1) Los modelos de dinámicas de crecimiento proveen un importante marco cuantitativo para caracterizar trayectorias epidémicas, generar estimaciones de parámetros claves, evaluar el impacto de las intervenciones de control, obtener información sobre la contribución de las diferentes vías de transmisión y producir pronósticos a corto y largo plazo. Estas ventajas motivaron la siguiente pregunta ¿se puede elegir el modelo de crecimiento más adecuado para una epidemia determinada?, la cual centró nuestro primer propósito que en el los **Capítulos** 1 y 2, intentan dar una luz sobre el desempeño de diferentes modelos de crecimiento en la descripción de brotes epidémicos reales. Específicamente, en el **Capítulo** 1, empleamos cuatro diferentes modelos de crecimiento basados en ecuaciones diferenciales (dos de estos con dos parámetros y dos con tres parámetros), y los exáminamos usando 37 bases de datos de diferentes brotes infeciosos que consisten en series de tiempo de casos de incidencia, para identificar en cada caso el mejor modelo para describir los crecimientos epidémicos. Por otro lado en el **Capítulo** 2 se intenta responder, ¿qué modelo de crecimiento epidémico es mejor para capturar las dinámicas generadas por otros modelos de crecimiento?, para ello se crea una metodología que ayuda a cuantificar las diferencias entre las dinámicas obtenidas de diferentes modelos que capturan procesos de crecimiento epidémico.

2) Dada la emergencia presentada por el COVID-19, y el conocimiento recopilado con los modelos fenomenológicos de crecimiento epidémico, surgen algunas inquietudes como, ¿para la enfermedad del COVID-19 los modelos generalizados captura mejor que los modelos simples? ¿el nuevo modelo sub-epidémico captura los múltiples peaks evidenciados por esta epidemia? ¿qué otros aportes nos puede entregar el ajuste de curvas de crecimiento epidémico, usando los ajustes con modelos fenomenológicos de crecimiento?, estas inquietudes nos llevó a trabajar junto a otros colegas, con quienes usamos datos de la epidemia presentada en Colombia y le aplicamos varios modelos fenomenológicos para comparar sus ajustes y pronósticos, y empleamos métricas de error entre los datos

y los ajustes de los modelos para determinar la calidad de los ajustes y sus pronósticos. Además el número efectivo de reproducción es calculado para entender el impacto de la epidemia evidenciado en Colombia. Tal desarrollo es presentado en el **Capítulo 3**.

3) Con respecto a los modelos epidemiológicos compartimentales y la situación vivida en Chile en el periodo inicial, donde las cuarentenas fueron aplicadas en diferentes unidades territoriales, con propósitos diferentes, nos preguntamos ¿qué modelo de ese tipo puede ayudar a modelar la situación presentada en Chile, que involucre las cuarentenas aplicadas, y nos permita medir su impacto? Es por ello que se propone un modelo compartimental homogeneamente mixto con las dinámicas de cuarentena para el caso chileno (combinando ideas de [31,81]).. También realizamos un estudio de identificabilidad para algunos conjuntos de parámetros a ajustar, los cuales deseamos que capturen las diferentes y particulares estrategias aplicadas en cada región chilena.

Contribuciones de esta tesis

La presente tesis se organiza como sigue:

En el **Capítulo 1**, para diferentes modelos fenomelógicos de crecimiento, proponemos un análisis comparativo entre cuatro modelos parámetricos basados en ecuaciones diferenciables ordinarias (ODEs), llamados modelo logístico y de Gompertz con sus respectivas generalizaciones que en cada caso consisten en elevar la función de incidencia acumulada a una potencia $p \in [0, 1]$. Este parámetro dentro de los modelos generalizados proporciona un criterio sobre el comportamiento del crecimiento temprano de la epidemia entre la incidencia constante para $p = 0$, crecimiento subexponencial para $0 < p < 1$ y el crecimiento exponencial para $p = 1$.

Los contenidos del **Capítulor 1** corresponden al artículo :

- R. Bürger, G. Chowell, and L. Y. Lara-Díaz. (2019). Comparative analysis of phenomenological growth models applied to epidemic outbreaks, *Mathematical Biosciences and Engineering*: MBE, 16(5), 4250–4273. `https://doi.org/10.3934/mbe.2019212`.

En el **Capítulo 2** contribuímos a un estudio sistemático de las diferencias entre modelos epidemiológicos y cómo tales diferencias pueden explicar la habilidad de ciertos modelos para proporcionar un mejor ajuste a los datos que otros. Para este fin, se define la medida de las distancias para describir las diferencias en las dinámicas entre diferentes modelos dinámicos. La distancia de un modelo de crecimiento a otro cuantifica qué tan bien se ajusta el primero a los datos generados por el segundo. Sin embargo, este concepto de distancia no es simétrico. El procedimiento de cálculo de distancias se aplica a datos sintéticos y a datos reales de brotes de influenza, ébola y COVID-19.

Los contenidos del **Capítulo 2** corresponden al artículo

- R. Bürger, G. Chowell, and L. Y. Lara-Díaz. (2021). Measuring differences between phenomenological growth models applied to epidemiology, *Mathematical Biosciences*, 334, 108558. `https://doi.org/10.1016/j.mbs.2021.108558`.

En el **Capítulo 3** se emplean diferentes modelos fenomenológicos de creciemiento epidémicos para modelar y caracterizar el brote de COVID-19 en Colombia, tales ajustes se realizan a nivel nacional y regional, de las cuales se logran varias estimaciones y curvas de ajuste, que luego permiten calcular el número efectivo de reproducción R_t, y hacer pronósticos a corto plazo. Este trabajo se hizo con el aporte de varios colegas, donde otros conceptos y cálculos son también incluídos.

Los contenidos del **Capítulo 3** corresponde al artículo [150]:

- A. Tariq, T. Chakhaia, S. Dahal, A. Ewing, X. Hua, S. K. Ofori, O. Prince, A. Salindri, A. E. Adeniyi, J. Banda, P. Skums, R. Luo, **L.Y. Lara-Díaz**, **R. Bürger**, I. C-H. Fung, E. Shim, A. Kirpich, A. Srivastava, **G. Chowell**. (2022). An investigation of spatial-temporal patterns and predictions of the coronavirus 2019 pandemic in Colombia, 2020–2021, *PLOS Neglected Tropical Diseases*, 16(3), e0010228. `https://doi.org/10.1371/journal.pntd.0010228`.

En el **Capítulo 4** introducimos una adaptación a una aproximación computacional para el estudio de la identificabilidad de los parámetros de un modelo compartimental, para este trabajo definimos un modelo para la propagación del COVID-19, inspirado en la situación de Chile a inicios del 2020, e incluimos una ruta de experimentación para verificar si un conjunto de parámetros estimados al ajustar una curva de datos con el modelo son identificables. Con esta idea se construye una metodología, donde datos sintéticos son construidos a partir del mismo modelo, y analizamos para un conjunto de parámetros de interés, qué tantos parámetros y combinaciones de estos el modelo es capaz de recuperar, variando también los tiempo de ajuste, con esta estrategia de combinar el número de parámetros a estimar y el tiempo de ajuste, desarrollamos una ruta para determinar los casos para los que se puede garantizar identificabilidad estructural y práctica. Al final, probamos para algunas regiones de Chile, los casos en que se puede concluir alguna de estas identificabilidades.

Los contenidos del **Capítulo 4** corresponde a la investigación:

- R. Bürger, G. Chowell, I. Kröker and L. Y. Lara-Díaz, Sensitivity and identifiability analysis for a model of the propagation and control of COVID-19 in Chile, *(en preparación)*.

CHAPTER 1

Comparative analysis of phenomenological growth models applied to epidemic outbreaks

This chapter shows a comparative study of various phenomenological growth models (PGMs) applied to 37 epidemic outbreaks to characterize the best model to capture the most growth patterns. The best model achieves the smallest error in terms of RMSE where the fit is compared to the real data.

1.1 Introduction

1.1.1 Scope

Dynamic growth models provide an important quantitative framework for characterizing epidemic trajectories, generating estimates of key transmission parameters, assessing the impact of control interventions, gaining insight to the contribution of different transmission pathways, and producing short- and long-term forecasts [32]. A natural question is that of the choice of the best suitable growth model for a given epidemic. It is the purpose of this paper to shed light on the performance of different growth models in describing different real epidemic outbreaks. Specifically, we employ four different growth models based on differential equations (two of them with two parameters, and two with three parameters), and apply them to a total of 37 infectious disease outbreak datasets consisting of time series of case incidence for different historic outbreaks comprising different diseases and settings.

The two-parameter models are the well-known logistic model (LM) [173] and Gompertz model (GoM) [77], and the three-parameter models are generalizations for both models which we refer to as the generalized logistic model (GLM) and the generalized Gompertz model (GGoM), respectively. These models incorporate a parameter p, which is an exponent that provides a criterion about the type of early growth dynamics, namely sub-exponential ($0 < p < 1$) or exponential ($p = 1$) growth. (For $p = 1$, the GLM and GGoM models reduce to the LM and

GoM models, respectively.) We explored the performance of these models in describing the trajectory of 37 outbreaks by applying the methodology described by Chowell [32] to estimate parameters with their confidence intervals. In this analysis, we analyzed how well models fitted the 37 outbreaks using the root mean squared error (RMSE).

The particular choice of parametric models complements that of [32], where the well-known exponential and Richards [127, 179] growth models are employed along with their generalized counterparts. Moreover, since that work is focused in detailing the methodology, the data set in [32] is limited to the 2013–2016 Ebola outbreak in Sierra Leone, and no mechanism of choice between two or more alternative models for the same data set is established. In this paper we are particularly interested in gaining insight into the types of outbreaks where the different model variants provide an enhanced description of the epidemic outbreaks.

1.1.2 Related work

This paper is focused on models given by ordinary differential equations (ODEs) to describe the temporal dynamics of epidemic outbreaks. The properties of ODEs as models of growth are treated in numerous monographs, see e.g. [13, 14, 17, 61, 82, 110, 141]. On the other hand, the presence of the nonlinearity caused by the growth rate exponent p precludes in some cases solutions of the corresponding ODE in closed form. Nevertheless, we mention that for $p = 1$, the properties of the Richards, logistic, Gompertz, and related (e.g., von Bertalanffy [174, 175]) models are broadly discussed in terms of closed algebraic expressions in [156–158] (see also the references cited in these papers).

We use phenomenological models within an empirical approach (without an explicit basis of physical laws or mechanisms) that are useful to reproduce the patterns observed in the time series data [170]. The result is a fairly simple temporal description of epidemic growth patterns [32]. For instance, epidemics display variable epidemic growth scaling (e.g., from sub-exponential to exponential). Here we are particularly interested in the contribution of the parameter p as a corrector in the fit and the possible improvement in the forecasts. The relevance of this parameter was recently highlighted by Chowell and Viboud [45] who demonstrated that a generalized-growth model is a simple tool that can be used to characterize the early epidemic growth profile from case incidence data as well as from synthetic data derived from transmission models via stochastic simulation [170]. Related references to early epidemic growth models also include [42, 94]. For the connection between the growth rate and the reproductive number of an epidemic, an aspect that is not discussed herein, we refer to [35, 48, 113, 178].

Finally, we mention that there are also stochastic models built to study sigmoidal behaviours. In particular, in recent years there have been many advances in stochastic models based on diffusion processes, particularly associated with the Gompertz and logistic curves. A general procedure for obtaining and estimating this type of models is considered in [130], where also further references can be found (see also [132]). As is discussed in the introduction of [130], considering particular choices of the time functions that define the exogenous factors has enabled researchers to define diffusion processes associated to alternative expressions of already-known growth curves [130, p. 2]. These processes include a Gompertz-type process [78] (applied to the study of rabbit growth), a generalized von Bertalanffy diffusion process (with an application to the growth of fish species) [129], a logistic-type process [131] (applied to the growth of a microorganism culture), and a Richards-type diffusion process [133]. More recent contributions to this line of research are [93] and [8].

1.2 Mathematical models

The general form of a phenomenological model is

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = f_i(x_1, \dots, x_n; \Theta), \quad i = 1, \dots, n, \tag{1.1}$$

where $\mathrm{d}x_i/\mathrm{d}t$ denotes the rate of change of the system state x_i, $i = 1, \dots, n$, and $\Theta = (\vartheta_1, \dots, \vartheta_m)$ is the set of model parameters, where the complexity of a model depends on the number m of parameters that are needed to characterize the states of the system and the spectrum of the dynamics that can be recovered from the model [32]. In this contribution we highlight the logistic growth model (LM) and the Gompertz model (GoM) and their respective generalizations, namely the generalized logistic model (GLM) and the generalized Gompertz model (GGoM). The last two models incorporate a parameter p that indicates the kind of scaling of growth. These models can be described as follows.

The logistic growth model (LM) relies on two parameters to characterize the trajectory of an epidemic, where the model is given by the differential equation

$$\frac{\mathrm{d}C}{\mathrm{d}t} = C'(t) = rC(t)\left(1 - \frac{C(t)}{K}\right), \tag{1.2}$$

where t is time, $C'(t)$ describes the incidence curve over time, $C(t)$ is the cumulative number of cases at time t, while the parameter $r > 0$ indicates the growth rate (its dimension is 1/time), and K is the size of the epidemic. During the initial stages of disease propagation, when $C(t) \ll K$, this model assumes an exponential growth phase, as can be inferred from the well-known explicit solution of (1.2),

$$C(t) = \frac{KC(0)\exp(rt)}{K + C(0)(\exp(rt) - 1)}.$$

The two-parameter Gompertz model (GoM) is given by the ODE

$$\frac{\mathrm{d}C}{\mathrm{d}t} = C'(t) = rC(t)\exp(-bt), \tag{1.3}$$

where the parameter $b > 0$ describes the exponential decay of the growth rate r, and the quantities C and C' have the same meaning as for the LM model. If $C(0)$ is the initial number of cases, then the solution of (1.3) is

$$C(t) = C(0)\exp\Big((r/b)\big(1-\exp(-bt)\big)\Big). \tag{1.4}$$

We generalize the logistic and Gompertz models by incorporating a growth scaling parameter $p \in [0,1]$ that indicates the kind of growth, where $p = 0$ corresponds to a constant incidence over time, $p = 1$ corresponds to the exponential growth and recovers the logistic model, and any value $0 < p < 1$ leads to a model that describes a sub-exponential growth, a property that leads to potentially more realistic models as shown in [170]. The model is given by the differential equation

$$\frac{\mathrm{d}C}{\mathrm{d}t} = C'(t) = rC^p(t)\left(1 - \frac{C(t)}{K}\right). \tag{1.5}$$

Similarly, the Gompertz model leads to the following ODE that defines the Generalized Gompertz Model (GGoM), where p plays the same role as in the GLM:

$$\frac{\mathrm{d}C}{\mathrm{d}t} = C'(t) = rC^p(t)\exp(-bt). \tag{1.6}$$

It is worth noting that for general values $p \in (0,1)$, (1.5) does not possess an explicit solution in closed algebraic form. (For a detailed discussion of this point and further references we refer to Ohnishi et al. [117], who deal with the Pütter-von Bertalanffy equation

$$\frac{\mathrm{d}C}{\mathrm{d}t} = \alpha C^A - \beta C^B$$

with positive constants α, β, A and B, which includes (1.5). Nevertheless, this equation admits an analytical solution given in implicit form [117, Eq. (9)].)

In contrast to the GLM equation (1.5), one may easily integrate the GGoM equation (1.6) for these values of p to get

$$C(t) = \left((1-p)\left(\frac{r}{b}\right)\big(1-\exp(-bt)\big) + C(0)^{1-p}\right)^{1/(1-p)}.$$

For this expression we get

$$C(t) \to \left(\frac{(1-p)r}{b} + C(0)^{1-p}\right)^{1/(1-p)} \quad \text{as } t \to \infty. \tag{1.7}$$

It is interesting to note that for the Gompertz model with $p = 1$, (1.3), the expression (1.4) implies that

$$C(t) \to C(0)\exp(r/b)$$

Growth model	Parameters
Logistic growth model (LM)	$\Theta = \{\vartheta_1 = r, \vartheta_2 = K\};\ r, K > 0$
Gompertz model (GoM)	$\Theta = \{\vartheta_1 = r, \vartheta_2 = b\};\ r, b > 0$
Generalized Logistic growth model (GLM)	$\Theta = \{\vartheta_1 = r, \vartheta_2 = p, \vartheta_3 = K\};\ r, K > 0,\ p \in [0,1]$
Generalized Gompertz model (GGoM)	$\Theta = \{\vartheta_1 = r, \vartheta_2 = b, \vartheta_3 = p\};\ r, b > 0,\ p \in [0,1]$

Table 1.1: Summary of information about models and parameters.

as $t \to \infty$ so the limit value depends linearly on $C(0)$ (unless the initial population is absorbed into b or r), while for $0 < p < 1$, (1.7) means that the limit of $C(t)$ still depends on $C(0)$ but does so in a nonlinear fashion.

Summarizing, we have two two-parameter models with their respective generalizations that are three-parameter models, where the third parameter is the growth scaling parameter $p \in [0,1]$, as we show in Table 2.1. Before we proceed, we illustrate by an example the effect of varying p within the GLM and GGoM, see Figure 3.2. We start with the logistic model (1.2) setting $r = 1$, $C(0) = 10$ and $K = 1000$. The solid red curve in Figure 3.2 (top left) shows the incidence curve $t \mapsto C'(t)$ corresponding to the solution $t \mapsto C(t)$ (Figure 3.2, top right). This solution approximates the maximum ($K = 1000$). Now we pass to the GLM (1.5) by gradually decreasing p from one to $p = 0.995$, $p = 0.99$, and so on (see the caption of Figure 3.2). We observe that the maxima of the incidence $C'(t)$ decrease (as follows easily from discussing the extrema of $C \mapsto C^p(1 - C/K)$), but their time of occurrence increases, as p is decreased. Furthermore, the incidence curves stay fairly close to the curve for $p = 1$ for values of p close to one, and all solutions behave like $C(t) \to K$ as $t \to \infty$.

In order to compare these observations with those for the Gompertz and GGoM models, we plot in Figure 3.2 (middle left) the incidence curve $t \mapsto C'(t)$ corresponding to the solution $t \mapsto C(t)$ (Figure 3.2, middle right) for the Gompertz model (1.3) with parameters $C(0) = 10$,

$$r = 1 - \frac{C(0)}{K} = 0.99, \quad \text{and} \quad b = \frac{r}{\ln\bigl(K/C(0)\bigr)} \approx 0.2150, \tag{1.8}$$

which have been chosen in such a way that $C'(0)$ is the same as for the GLM as well as that $C(t) \to K$ as $t \to \infty$ (cf. (1.4)) for $p = 1$. Note that the maximum of $C'(t)$ is smaller than for the logistic model. As p is decreased, but all other parameters are kept, these maxima become smaller (as with the GLM), but they appear each time *earlier* (in contrast to the GLM). However, for $t \to \infty$ we observe that consistently with (1.7), $C(t)$ approaches smaller values than K as $t \to \infty$. If we wish to ensure that the GGoM with $p \in (0,1)$ has the same value of $C'(0)$ as the GLM (for the corresponding value of p) *and* $C(t) \to K$ as $t \to \infty$, then we must also adjust b by setting

$$r = 1 - \frac{C(0)}{K}, \quad b = \frac{r(1-p)}{\bigl(K^{1-p} - C(0)^{1-p}\bigr)} \tag{1.9}$$

(which results from equating the limit in (1.7) with K). From the bottom plots of Figure 3.2 we observe that the joint variation of p and b produces curves similar to those of the GLM.

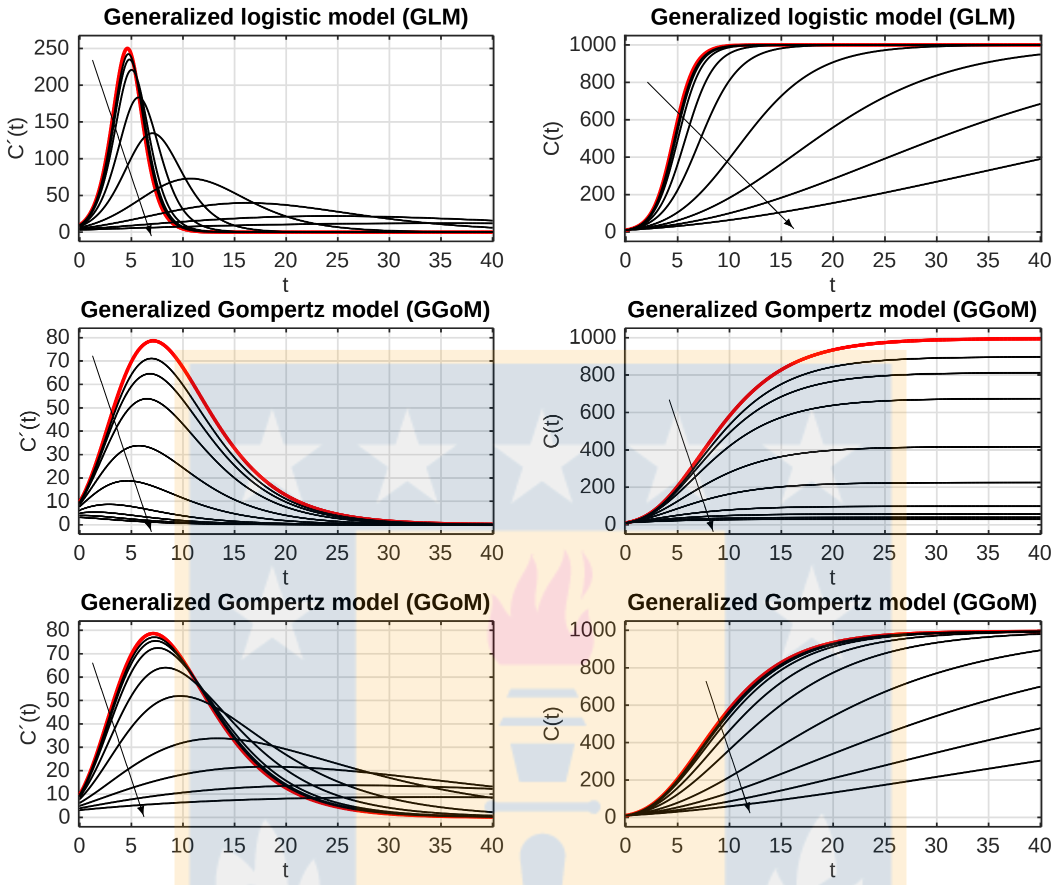


Figure 1.1: Illustration of the GLM model (top) and the GGoM model (middle and bottom), showing in each case $C'(t)$ (left) and $C(t)$ (right). The solid red curve corresponds to $p = 1$. The arrow indicates decreasing values of $p = 0.995$, 0.99, 0.98, 0.95, 0.9, 0.8, 0.7, 0.6, and 0.5, corresponding to the thin black curves. The plots in the middle correspond to fixed values of r and b (see (1.8)), while in the bottom r is fixed but b is variable (see (1.9)).

Finally, let us emphasize once again that the exponent p is introduced in both (1.5) and (1.6) in such a way that it affects the *initial* growth rate, corresponding to the early stage when $C(t)/K \ll 1$ and therefore $C'(t) \approx rC^p(t)$, so that p characterizes sub-exponential growth dynamics [170]. In particular, the identification of p at early stage of an epidemic is fundamental for forecasting the outbreak [45]. It is therefore instructive to provide an example to compare (1.5) with an alternative way of introducing an exponent p into (1.2), namely the well-known Richards equation [127]

$$\frac{\mathrm{d}C}{\mathrm{d}t} = C'(t) = rC(t)\left(1 - \left(\frac{C(t)}{K}\right)^p\right) = \left(\frac{r}{K^p}\right) C(t)\bigl(K^p - C(t)^p\bigr). \tag{1.10}$$

Figure 1.2 displays the incidence curves $t \mapsto C'(t)$ and the solution $t \mapsto C(t)$ for selected values of p for both the GLM model (1.5) and the Richards equation (1.10). We observe that

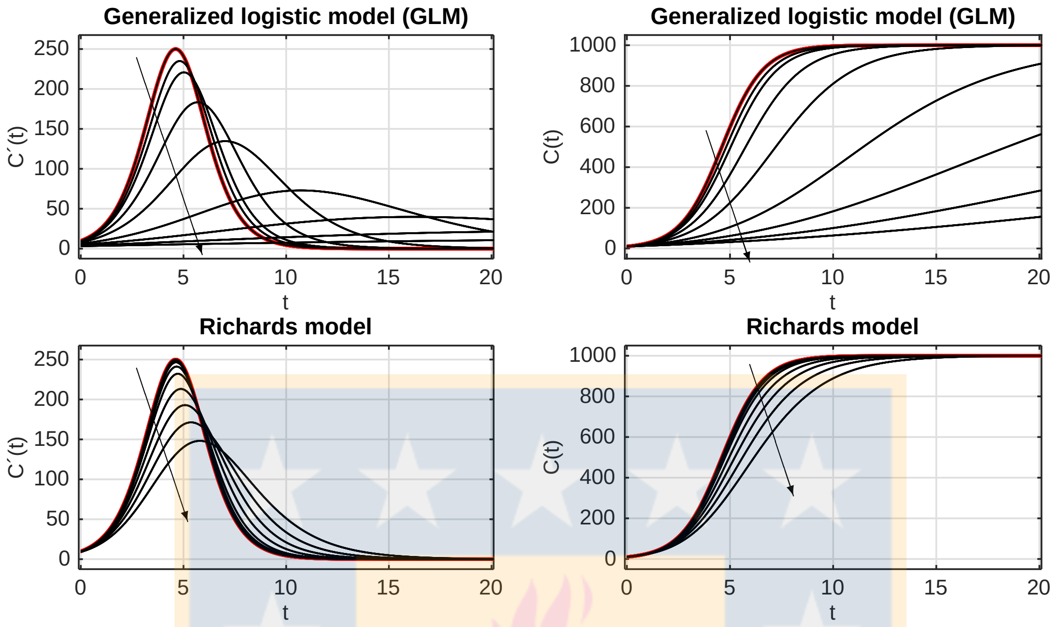


Figure 1.2: Illustration of the GLM model (top) and the Richards model (bottom), showing in each case $C'(t)$ (left) and $C(t)$ (right), starting from $C(0) = 10$ with $K = 1000$. The solid red curve corresponds to $p = 1$. The arrow indicates decreasing values of $p = 0.99$, 0.98, 0.95, 0.9, 0.8, 0.7, 0.6, and 0.5, corresponding to the thin black curves.

since $C(0)/K \ll 1$, the initial growth rates for (1.10) are very similar for all values of p, in contrast to those of the GLM model. Thus, the variability of the exponent p in the Richards equation (1.10) is not suitable for capturing sub-exponential initial growth.

On a similar note, we mention that the traditional form of the Gompertz ODE (cf., e.g., [78]) is

$$\frac{\mathrm{d}C}{\mathrm{d}t} = C'(t) = \alpha \ln \left(\frac{K}{C(t)} \right) C(t) = (\alpha \ln K) C(t) - \alpha C(t) \ln C(t) \tag{1.11}$$

with a constant $\alpha > 0$, which is a nonlinear differential equation, in contrast to the linear ODE (1.3) utilized herein. Our preference of (1.3) is based on the fact that this equation can easily be equipped with the exponent p to give (1.6). Furthermore it is fairly easily possible to compare (1.6) and its solutions with those of the sub-exponential growth equation

$$\frac{\mathrm{d}C}{\mathrm{d}t} = rC(t)^p$$

analyzed in [45, 170], while the multiple, and nonlinear occurrence of $C(t)$ makes such a generalization at least more complicated.

Case No.	Disease	Outbreak	Temporal resolution	Total data	Case No.	Disease	Outbreak	Temporal resolution	Total data
1	Ebola	Forecariah (GIN)	weeks	51	20	Ebola	Tonkolili (SLE)	weeks	29
2	Ebola	Gueckedou (GIN)	weeks	49	21	Ebola	Western Area Rural (SLE)	weeks	51
3	Ebola	Keroune (GIN)	weeks	14	22	Ebola	Western Area Urban (SLE))	weeks	55
4	Ebola	Kindia (GIN)	weeks	30	23	Ebola	Grand Bassa (LBR)	weeks	30
5	Ebola	Macenta (GIN)	weeks	32	24	Ebola	Congo (1976)	days	52
6	Ebola	N'Zerekore (GIN)	weeks	24	25	Ebola	Uganda (2000)	weeks	18
7	Ebola	Bomi (LBR)	weeks	33	26	Measles	London (ING) (1948)	weeks	40
8	Ebola	Bong (LBR)	weeks	17	27	Plague	Bombay (IND) (1905-06)	weeks	41
9	Ebola	Grand Cape Mount (LBR)	weeks	29	28	Plague	Madagascar (2017)	weeks	50
10	Ebola	Lofa (LBR)	weeks	24	29	Smallpox	Khulna (BGD) (1972)	weeks	13
11	Ebola	Margibi (LBR)	weeks	40	30	Yellow fever	Luanda (AGO) (2016)	weeks	28
12	Ebola	Montserrado (LBR)	weeks	42	31	FMD	UK (2001)	days	121
13	Ebola	Bo (SLE) (2014)	weeks	39	32	FMD	Uruguay (2001)	days	27
14	Ebola	Kailahun (SLE)	weeks	33	33	Pandemic Influenza	San Francisco (USA) (1918)	days	63
15	Ebola	Kambia (SLE)	weeks	45	34	Zika	Antioquia (COL)(2016)	days	105
16	Ebola	Kenema (SLE)	weeks	39	35	VIH-AIDS	Japan (1985-2012)	years	21
17	Ebola	Kono (SLE)	weeks	30	36	VIH-AIDS	NYC (1982-2002)	years	70
18	Ebola	Moyamba (SLE)	weeks	37	37	Cholera	Aalborg (DNK) (1853)	days	105
19	Ebola	Port Loko (SLE) (2014)	weeks	54					

Table 1.2: Information on the 37 data sets of epidemic outbreaks obtained from the following sources: Cases 1 to 23: [118], Case 24: [25, 67], Case 25: [38, 182], Case 26: [11], Case 27: [3], Case 28: [184], Case 29: [144], Case 30: [183] , Cases 31 and 32: [40, 41], Case 33: [39], Case 34: [36], Cases 35 and 36: [73, 79], Case 37: [121].

1.3 Materials and methods

In order to compare the mathematical models, we need time series data that describe the temporal changes in one or more states of the system, whose temporal resolution varies among daily, weekly or yearly and by the frequency at which the state of the system is measured. We herein employ a data set for 37 different epidemic trajectories with different temporal resolutions (see Table 1.2). Additionally we present the method for fitting the model to the data, that is, to estimate the parameters as in [32]. Finally, to compare the models, we conduct a comparative analysis of RMSEs for all models and epidemics. Then, to continue we present the materials and methods that allow us to understand the methodology.

1.3.1 Datasets

Table 1.2 summarizes the information of the 37 epidemic outbreaks analyzed, including the name of the disease associated with each epidemic, the location where the outbreak occurred, the temporal resolution (by days, weeks, or years) of the time series, and the number of data points. For each outbreak, the onset corresponds to the first observation associated with a monotonic increase in incident cases, up to the peak incidence. We notice that for Ebola we have more information about the outbreak in West Africa (see also [46, 47, 119]).

1.3.2 The root mean square error (RMSE)

As in [32], besides using the residuals for any systematic deviations for the model fit to the data, it is also possible to quantify the error of the model fit to the data using performance metrics [88]. These metrics are also useful to quantify the error associated with a forecast. A widely used performance metric is the root mean squared error (RMSE) given by

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(f(t_i,\hat{\Theta}) - y_{t_i}\right)^2},$$

where $\hat{\Theta}$ is the set of parameter estimates, $f(t_i,\hat{\Theta})$ denotes the best-fit model, and y_{t_i} $(i = 1,\ldots,n)$ is the time series data (for that specific epidemic outbreak) and n is the total number of data points. In this work we employ the RMSE since this quantity naturally arises in the context of least-squares methods. Other applicable performance metrics [32] include the mean absolute error (MAE) and the mean absolute percentage error (MAPE), given by the respective expressions

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|f(t_i,\hat{\Theta}) - y_{t_i}\right|, \quad \mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{f(t_i,\hat{\Theta}) - y_{t_i}}{y_{t_i}}\right|.$$

While we have not applied any special treatment on outliers when calculating the RSME, the sensitivity of each of these performance metrics to anomalous data is left as a topic for future study.

1.3.3 Parameter estimation and confidence interval generation

Based on the description of the determination of the best fit in [32], we use the built-in Matlab (The Mathworks, Inc.) function LSQCURVEFIT to obtain parameter estimates via least-square fitting of the model solution to the observed data. This is achieved by searching for the set of parameters $\hat{\Theta} = (\hat{\vartheta}_1,\ldots,\hat{\vartheta}_m)$ that minimizes the sum of squared differences between the observed data $y_{t_i} = y_{t_1},\ldots,y_{t_n}$ and the corresponding model solution denoted by $f(t_i,\Theta)$. For the implementation for this function, we need the initial parameter guesses and the upper and lower bounds for these parameters as well as the initial data point C(0) . The process for the parameter estimation is summarized in the next steps:

1. Define the upper and lower bounds for each parameter.

2. Consider m sets of initial parameters defined with the Matlab function LSHDESING and the upper and lower bounds defined in step 1.

3. Calculate the parameter estimation for each set of initial parameters with the function LSQCURVEFIT.

4. Measure the error RMSE and select the parameter estimates with lowest RMSE, in order to ensure that the global minimum rather than a local minimum was found.

On the other hand, to generate the confidence interval, we use the parametric bootstrap method [66] (see also [34, 43]) with Poisson error structure that was implemented to generate 250 model realizations. This process can be summarized in the following steps:

1. With the parameter estimations $\hat{\Theta}$ obtained by the least-squares fit of the model $f(t_i, \Theta)$ to the time series data $y_{t_1}, \dots, y_{tn}$, we achieve the best-fit model $f(t_i, \hat{\Theta})$.

2. Then, we generate S-times replicated simulated datasets, using the best-fit model, which we denote by $f_1^*(t_j, \hat{\Theta}), \dots, f_S^*(t_j, \hat{\Theta})$. To generate these simulated data sets, we first use the best-fit model $f(t_i, \hat{\Theta})$ to calculate the corresponding cumulative curve function $F(t_j, \hat{\Theta})$ defined as

$$F(t_j, \hat{\Theta}) = \sum_{l=1}^{j} f(t_l, \hat{\Theta}), \quad j = 2, \dots, n.$$

   Moreover, $f_k^*(t_1, \hat{\Theta}) = f(t_1, \hat{\Theta})$ for $k = 1, \dots, S$. Besides, these data are generated assuming a Poisson error structure as follows: we assume that

$$f_k^*(t_j, \hat{\Theta}) = \mathrm{Po}\big(F(t_j, \hat{\Theta}) - F(t_{j-1}, \hat{\Theta})\big), \quad j = 2, 3, \dots, n, \quad k = 1, 2, \dots, S,$$

   where $\mathrm{Po}(\lambda)$ denotes the Poisson distribution with mean λ.

3. We re-estimate parameters for each of the S simulated realizations, which are denoted by $\hat{\Theta}_i$ for $i = 1, \dots, S$.

4. Finally, using the set of re-estimated parameters $\hat{\Theta}_i$, $i = 1, \dots, S$, we construct the confidence interval, so the resulting uncertainty around the model fit is given by

$$f(t, \hat{\Theta}_1), f(t, \hat{\Theta}_2), \dots, f(t, \hat{\Theta}_S).$$

Then, for our case, from these $S = 250$ realizations, we calculate 95% confidence intervals for model parameters.

1.3.4 Methodology: Analysis of the RMSE

In this section we summarize the methodology used to decide which is the best model for a given outbreak, and to analyze the contribution of the parameter p. The definitions and theory are taken from [32]. The methodology consists in an analysis of the RMSE error with the help of bar and scatter charts.

For this purpose, we first explore the initial parameters for each model and epidemic in order to ensure that the best fit of the model yields the smallest RMSE following the steps defined in the Section 3.3 for parameter estimation and considering $r, b \in [0, 5]$, $K \in [0, 10^7]$ and the known $p \in [0, 1]$. The above is an important process in order to ensure that we are obtaining the best fit to the data using the LSQCURVEFIT function in Matlab. We then with the best fits for each model and epidemic, we have their incidence curves and the lower RMSE. With these values we obtain graphs that compare the fit with the data, bar charts and scatter plots, which will be used for the error analysis (see Figure 1.3).

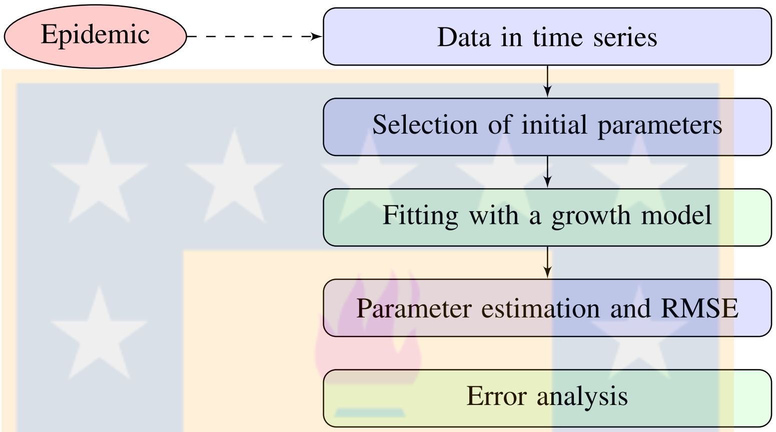

Epidemic
Data in time series
Selection of initial parameters
Fitting with a growth model
Parameter estimation and RMSE
Error analysis

Figure 1.3: Methodology for error analysis.

1.4 Results

1.4.1 Error analysis and comparison of fits for each epidemic

With the RMSE and the best fits obtained for each model, we obtain tables and graphics (see Table 1.3 and Figures 1.4 to 1.8) to compare the sizes of the errors for each model and epidemic outbreak, where the numbers from 1 to 37 in Table 1.3 identify the cases of outbreak (see Table 1.2). In Table 1.3 we observe that (independently of the epidemic) the GLM method yields the lowest RMSE in most of the cases (highlighted in green), and the LM yields the larger errors (highlighted in yellow). Besides, whenever the GLM is not the “best” model, the GGoM follows.

Furthermore, we also observe that between LM and GoM, the GoM is better, because the dynamics of this model are more closely aligned to the dynamics of the GGoM. Furthermore, the LM is associated with the largest errors in the great majority of the cases of outbreaks.

Case No.	LM	GLM	GoM	GGoM	Case No.	LM	GLM	GoM	GGoM
1	5.91840	5.08749	5.28578	5.28578	20	13.81574	10.44249	10.33338	10.33338
2	5.36334	4.68065	4.72246	4.71404	21	22.13303	12.51529	13.31396	13.31396
3	6.80378	5.21596	5.21959	5.19461	22	27.16583	20.98778	26.48861	26.48861
4	3.07198	3.06934	3.22175	3.22175	23	3.01595	2.47452	2.57271	2.43990
5	16.02456	8.74242	8.73707	8.49820	24	3.03925	2.28213	2.29532	2.29498
6	6.95680	5.27913	5.36135	5.36135	25	9.36028	6.37029	7.81157	7.81157
7	5.42450	3.96942	4.41215	3.96139	26	264.91368	108.36306	147.87904	147.87904
8	7.01087	5.81503	6.21805	5.81937	27	57.27638	51.60129	154.36235	154.36234
9	4.79101	4.79101	5.05457	5.05457	28	20.21720	8.50542	8.31521	7.87152
10	8.58955	8.58955	14.88488	14.88488	29	31.10051	28.45452	31.44816	31.44816
11	14.13951	11.40156	17.78045	17.78045	30	16.22091	9.42127	13.00660	13.00660
12	22.89522	14.77254	37.63692	37.63692	31	7.59491	5.12274	5.79428	5.79428
13	19.73810	10.08899	12.70424	12.70424	32	265.53459	78.47628	118.53622	79.95863
14	17.94184	11.77214	12.98507	11.93256	33	137.38697	137.38697	387.23469	387.23464
15	4.13574	3.31649	3.35153	3.34541	34	10.15666	5.47679	5.54259	5.54259
16	9.18180	5.58002	5.76447	5.74384	35	2174.08795	1354.63027	1493.07521	1493.07521
17	13.74655	13.74655	17.83847	17.83847	36	11.13642	7.40159	8.17479	7.64371
18	11.77779	11.32307	11.31585	11.31585	37	31.65064	26.58298	46.71374	46.71374
19	26.11925	11.66119	12.71813	12.71813					

Table 1.3: RMSE using the total data for each model. For each outbreak, we highlight the lowest RMSE (green) and the highest value (yellow) for the error sizes.

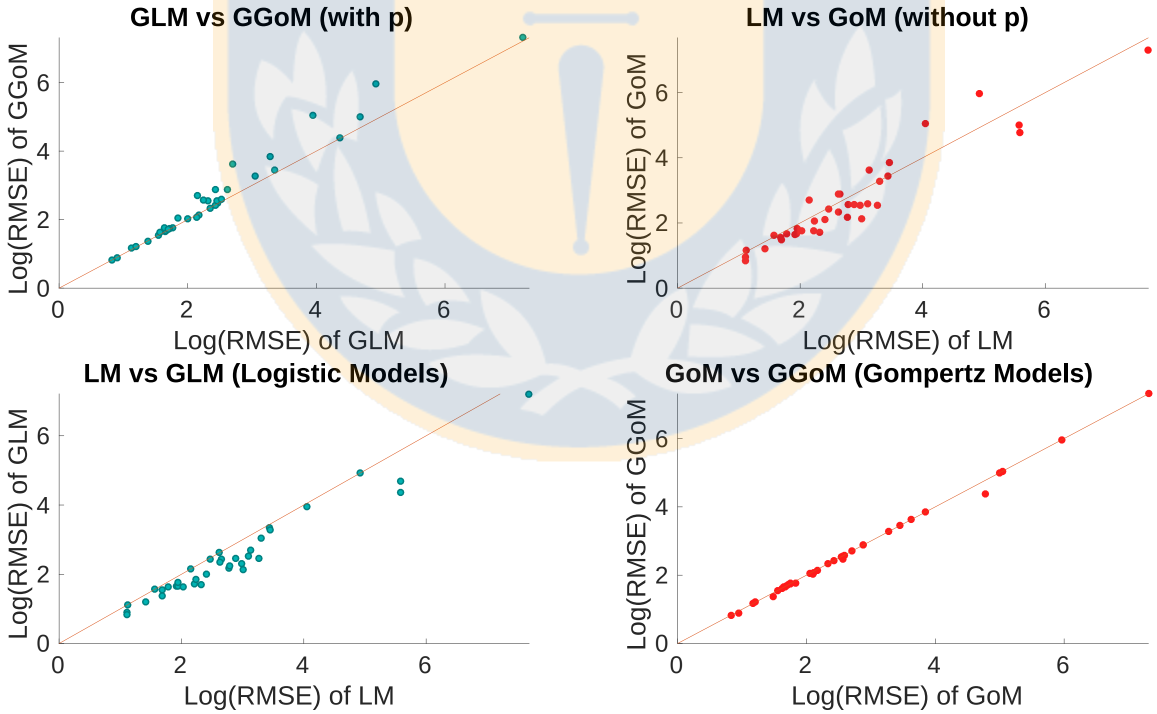


Figure 1.5: Scatter plots for RMSE, where we verify that the pair of Gompertz models have a closer behavior than the logistic models, where the variations are more marked. Additionally, we also verify that the models incorporating the parameter p yield similar errors, in contrast to the models with $p = 1$.

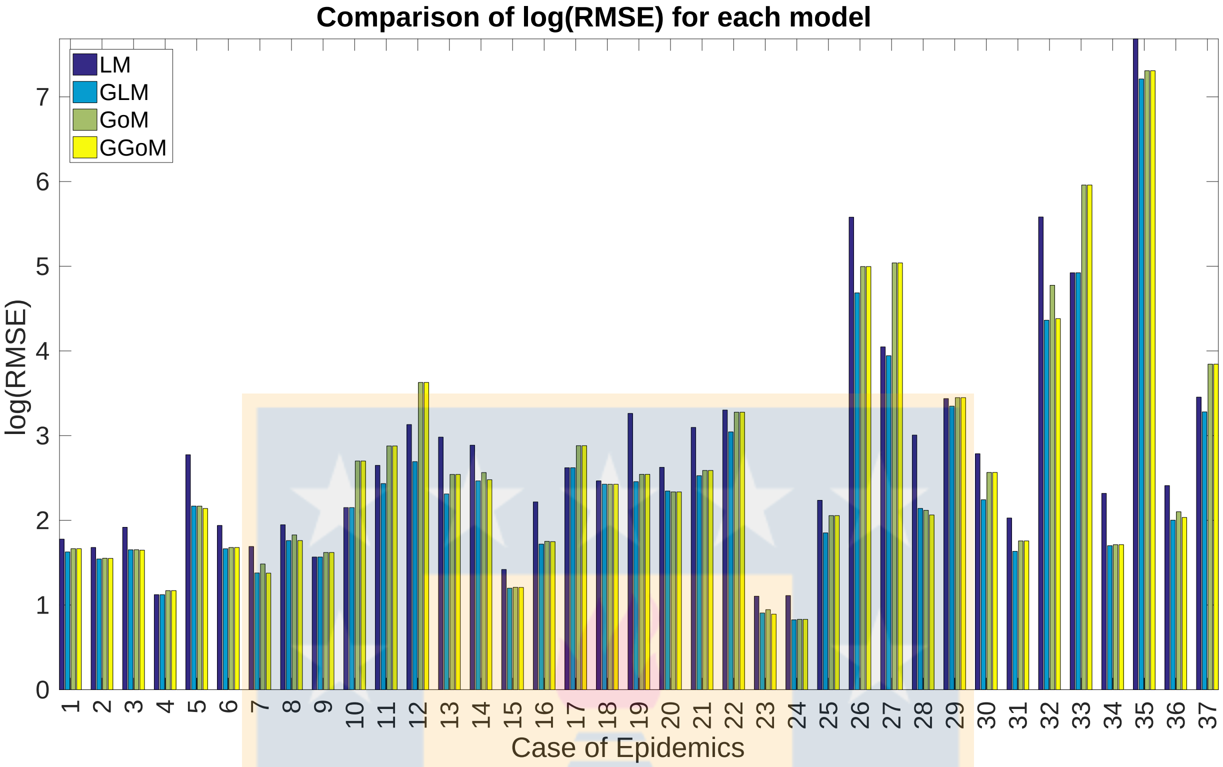


Figure 1.4: Bar chart for comparison of the errors of each methods, where we observed that among the best results are for the GLM and GGoM.

Figures 1.4 and 1.5 display the RMSE for each model and dataset. In Figure 1.4 we can see that although the GLM outperforms in most cases, we note that the error for the GLM is higher for Cases 3, 5, 7, 18, 20, 23, and 28 compared to the GGoM. Yet, those error differences are very small.

We also employ scatter plots to compare the errors yielded by a pair of models across all of the epidemics (Figure 1.5). Therefore, we compare the models with or without the parameter p, and then between the logistic and Gompertz models. For the first comparison we verify that the GGoM has errors with sizes larger than the GLM, unlike the models without p, where the behavior is different, since the LM has the errors with more scatter and below the line with slope one. Moreover, for the second group of cases, we note that the logistic models have a more scattered behavior above the diagonal line, where LM has errors with sizes greater than the sizes for the GLM's errors. This contrasts with the Gompertz models, where the scatter is closer to the diagonal. This shows that the errors yielded by both Gompertz models are very similar, and we can readily observe that these models are stable or closer to each other.

Having analyzed the RMSE for each model, now we study their respective fits for each epidemic outbreak, where we obtain a graphic sample of the best fit that corresponds to the RMSE, i.e., we will plot the best fits. These results are plotted in Figures 1.6 to 1.8. In these figures we can observe and compare the quality of the fits and their erorrs, where can note that the best fits to the data correspond to the smaller errors in terms of the RMSE.

Having finalized our comparative analysis of the model fits and their corresponding errors, we point out that for the Ebola epidemics (Cases 1 to 25), the GLM tends to yield an improved description to the data because in those cases where the GGoM wins (in terms of smallness of the RMSE), the corresponding errors do not differ by more than 0.6399. However, for the rest of the cases of epidemic outbreaks, the best model remains the GLM which yields smaller errors compared to the GGoM.

1.4.2 Parameter estimation

These results were obtained from the fits calculated in the previous section with the use of the Matlab function LSDCURVEFIT. We summarize the results for all cases in Table 1.4. We note that for the GGoM, there are 24 cases with $p = 1$, which means that these exhibit an initial exponential growth, where moreover the Gompertz and GGoM models yield equal RSMEs for that value of p. On the other hand for this same period of time and for the logistic models, we notice that only for four epidemics we have $p = 1$ (exponential initial growth), and the others give rise to initial sub-exponential growth with $p \in (0, 1)$. There were a number of outbreaks where the Gompertz models yield $p = 1$ (Gompertz and GGoM models are equivalent), for which the differences between the corresponding RMSEs are negligible.

Additionally, we observe that for the cases of Ebola in Grand Cape Mount, Lofa, Kono and Pandemic Influenza (Cases 9, 10, 17, and 33), we obtained $p = 1$ for the two generalized models. Also, for epidemics when the value of p for GLM is near one, the corresponding value of the parameter for the GGoM is one including the epidemics of Ebola (Kindia, Montserrado; Cases 4 and 12), Plague (Bombay; Case 27) and Cholera (Aalborg; Case 37), in Table 1.4. We also observe that when the value of the parameter for GLM is small, for example the cases of Ebola (Bomi, Bong; Cases 7 and 8), the value for the GGoM is also small, and for all cases when the value of $p = 1$ en GGoM, the values of p for GLM is greater than 0.6.

1.4.3 Confidence intervals

In this part, for the calculation of confidence intervals, we consider the generalized models (GLM and GGoM), for which we can obtain another piece of information to compare both models, and to decide which models best fit a given dataset. To this end we take the same initial parameters obtained for the RMSE calculation, and we use the parametric bootstrap process with 250 simulations with Poisson error structure, defined in Section 1.3, and summarize the results in Tables 1.5 and 1.6. In these results we note that the intervals are narrower and contain the mean value, suggesting that the parameters are identifiable (see [32]) for the GLM model. On the other hand, for the GGoM model, this situation occurs in some cases, for example, see

Figure 1.9, where for Case 1 the confidence interval obtained with GLM model has a bar chart that is centred, while that for the GGoM model, the bar chart displays a distribution with two modes. This behavior displayed by the GGoM model can be due to dependency or correlations (presented in Section 1.2) between the parameters b and p.

Another observation is that the non-identifiability can be present in the results where the upper and lower limit of the 95%CI intervals are not so close, and the mean is not a central value inside the interval. This is observed for the GGoM in the Cases 1 and 24, and the opposite situation can be observed, for instance, for Cases 12 and 19, where the mean value is a central value inside the interval which has the extremes very close. This last situation also appears in all the results derived from the GLM.

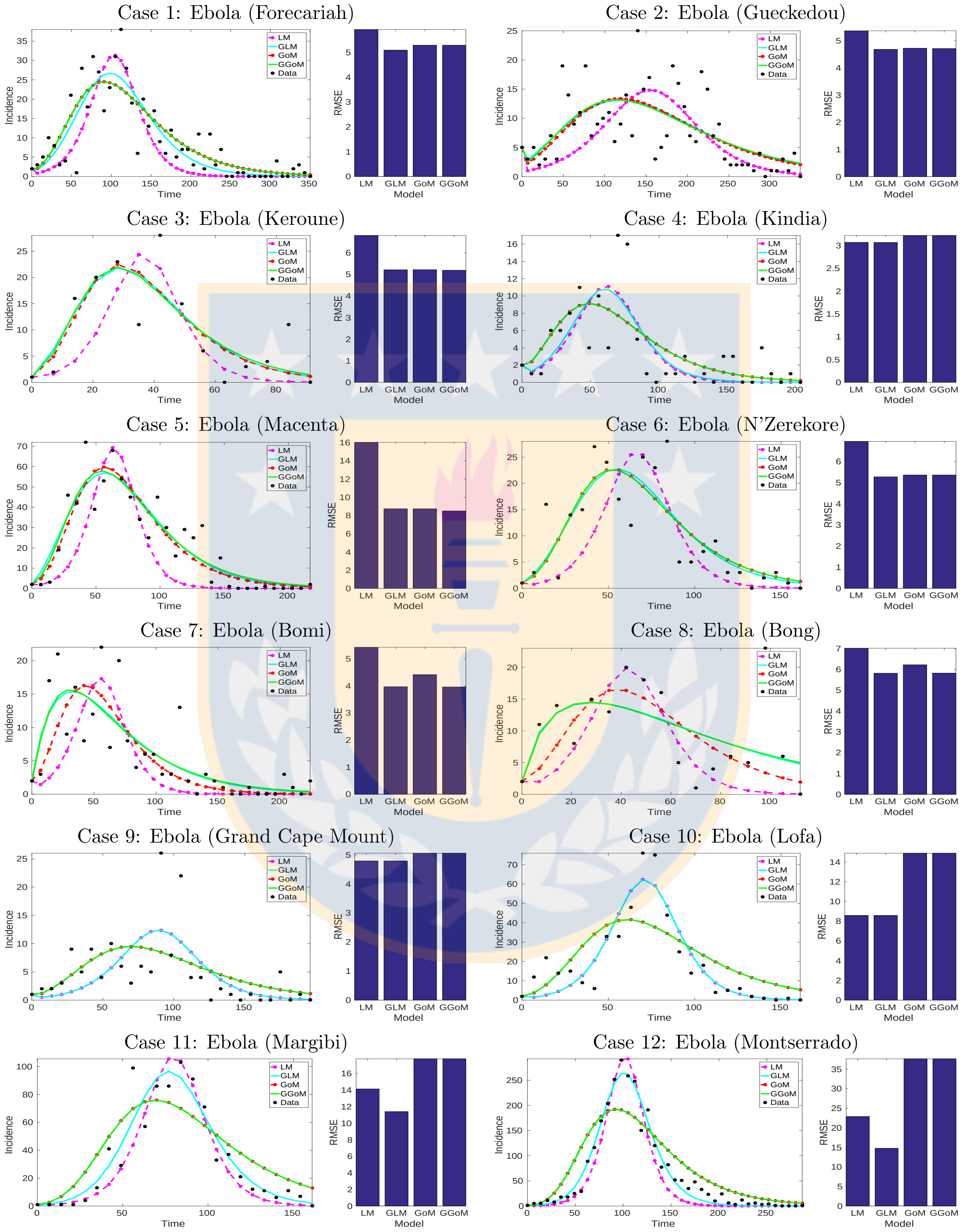
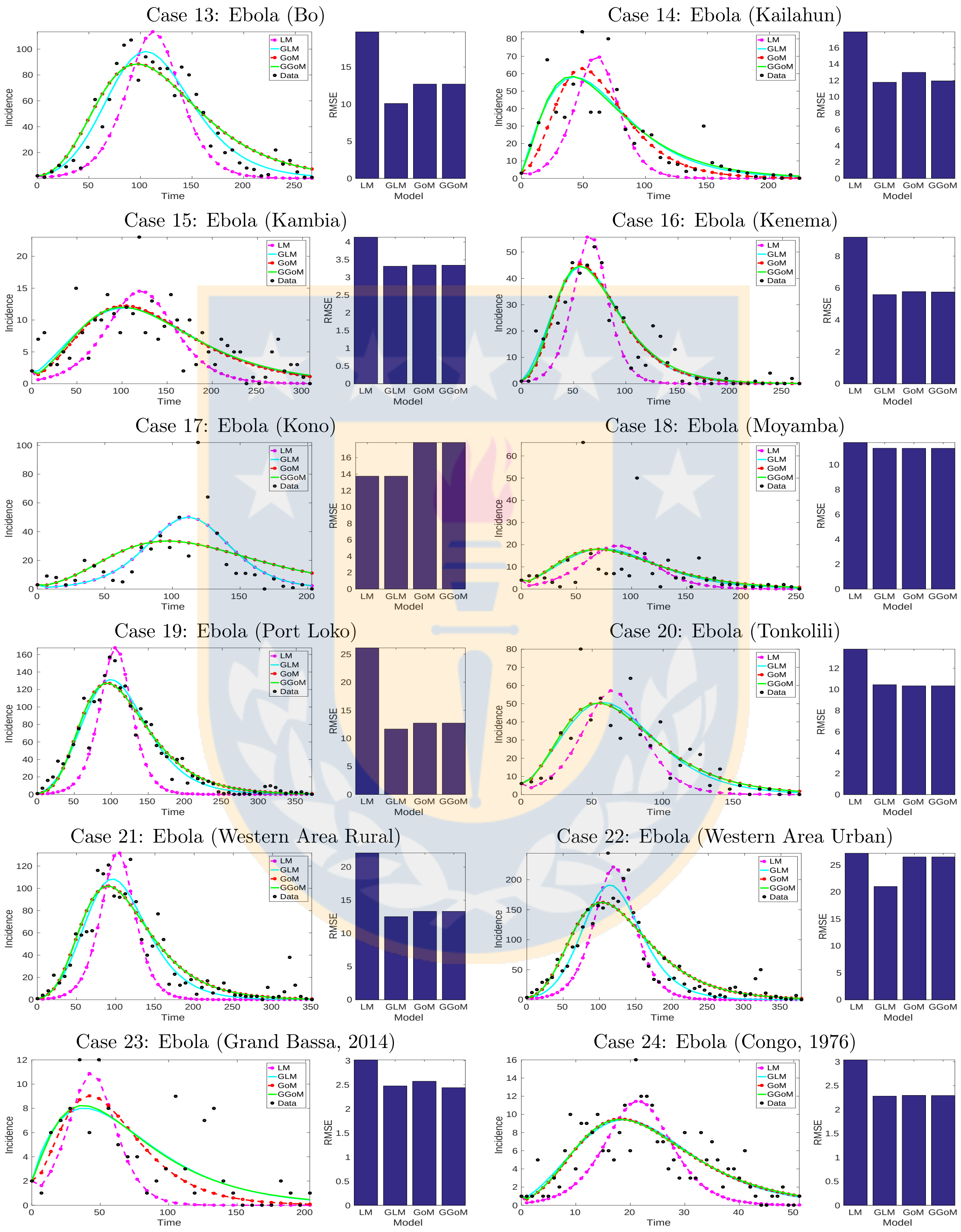

Case 1: Ebola (Forecariah)
Case 2: Ebola (Gueckedou)
Case 3: Ebola (Keroune)
Case 4: Ebola (Kindia)
Case 5: Ebola (Macenta)
Case 6: Ebola (N'Zerekore)
Case 7: Ebola (Bomi)
Case 8: Ebola (Bong)
Case 9: Ebola (Grand Cape Mount)
Case 10: Ebola (Lofa)
Case 11: Ebola (Margibi)
Case 12: Ebola (Montserrado)

Figure 1.6: Results of fits for epidemic outbreaks (Cases 1 to 12).



Case 13: Ebola (Bo)

Case 14: Ebola (Kailahun)

Case 15: Ebola (Kambia)

Case 16: Ebola (Kenema)

Case 17: Ebola (Kono)

Case 18: Ebola (Moyamba)

Case 19: Ebola (Port Loko)

Case 20: Ebola (Tonkolili)

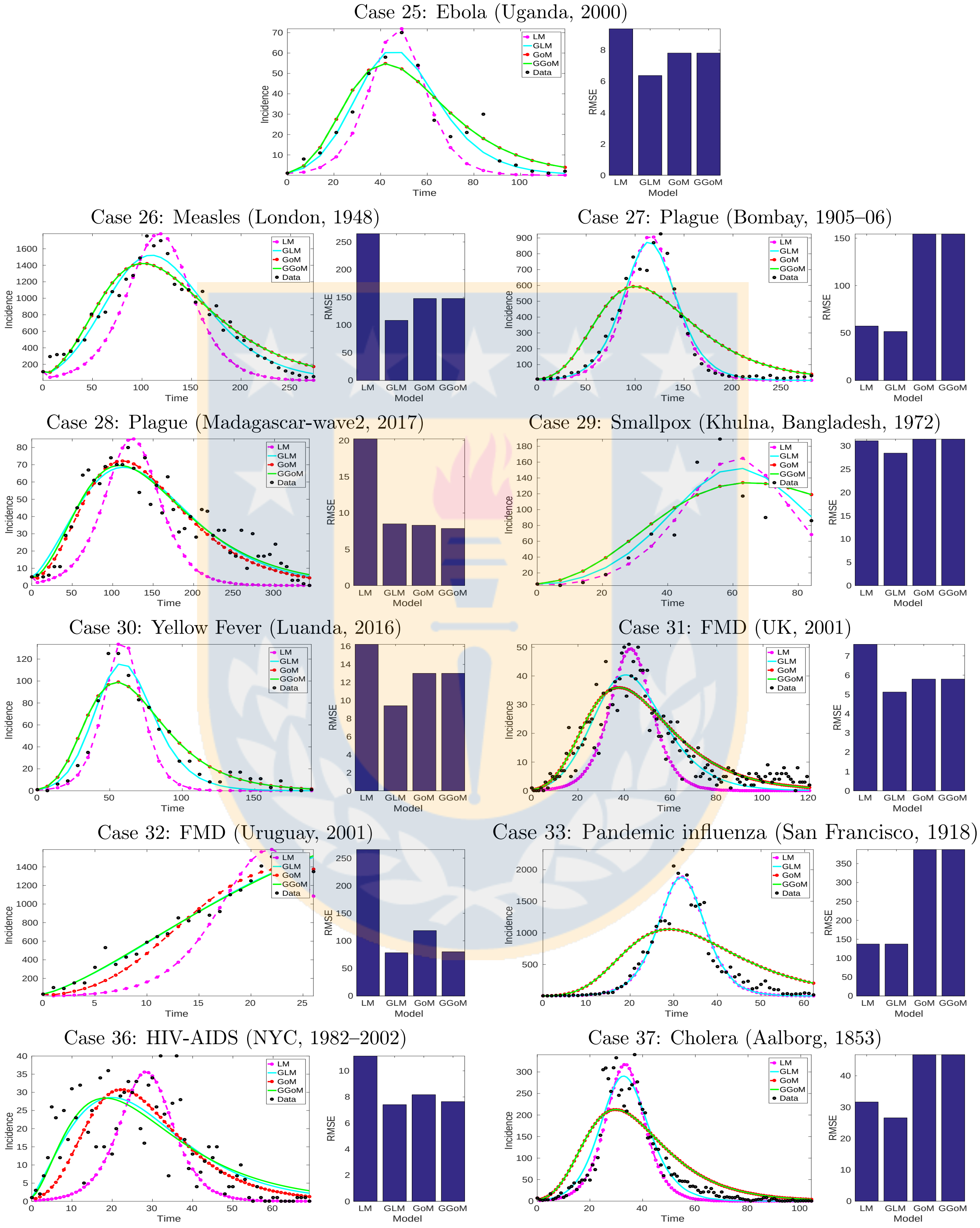
Case 21: Ebola (Western Area Rural)

Case 22: Ebola (Western Area Urban)

Case 23: Ebola (Grand Bassa, 2014)

Case 24: Ebola (Congo, 1976)

Figure 1.7: Results of fits for epidemic outbreaks (Cases 13 to 24).



Case 25: Ebola (Uganda, 2000)

Case 26: Measles (London, 1948)

Case 27: Plague (Bombay, 1905–06)

Case 28: Plague (Madagascar-wave2, 2017)

Case 29: Smallpox (Khulna, Bangladesh, 1972)

Case 30: Yellow Fever (Luanda, 2016)

Case 31: FMD (UK, 2001)

Case 32: FMD (Uruguay, 2001)

Case 33: Pandemic influenza (San Francisco, 1918)

Case 36: HIV-AIDS (NYC, 1982–2002)

Case 37: Cholera (Aalborg, 1853)

Figure 1.8: Results of fits for epidemic outbreaks (Cases 25 to 37).

Case	LM		GLM			GoM		GGoM		
no.	$\hat{r}$	$\hat{K}$	$\hat{r}$	$\hat{p}$	$\hat{K}$	$\hat{r}$	$\hat{b}$	$\hat{r}$	$\hat{b}$	$\hat{p}$
1	0.0515	349.7231	0.1166	0.7713	444.2758	0.1074	0.0195	0.1074	0.0195	1.0000
2	0.0272	312.0756	0.1333	0.6196	407.0675	0.0716	0.0129	0.0716	0.0119	0.9199
3	0.1395	103.0945	0.4286	0.6217	140.4582	0.3545	0.0636	0.3545	0.0580	0.9182
4	0.0661	96.5409	0.0761	0.9502	99.1794	0.1245	0.0307	0.1245	0.0307	1.0000
5	0.0895	447.7870	0.4231	0.6114	725.6594	0.2533	0.0334	0.2533	0.0293	0.8988
6	0.0824	181.5802	0.2410	0.6776	248.1927	0.1937	0.0350	0.1937	0.0350	1.0000
7	0.0794	125.5729	0.6590	0.3822	197.6517	0.5303	0.0376	0.5303	0.0229	0.5241
8	0.1022	112.4554	0.8366	0.3050	197.8649	0.7161	0.0421	0.7161	0.0188	0.3987
9	0.0563	126.3540	0.0563	1.0000	126.3550	0.1222	0.0243	0.1222	0.0243	1.0000
10	0.0801	449.7942	0.0801	1.0000	449.7945	0.1680	0.0300	0.1680	0.0300	1.0000
11	0.0860	717.8667	0.1321	0.8856	835.4420	0.2037	0.0295	0.2037	0.0295	1.0000
12	0.0781	2186.2506	0.1151	0.9075	2558.8591	0.1891	0.0234	0.1891	0.0234	1.0000
13	0.0580	1120.3306	0.1510	0.7861	1516.9220	0.1379	0.0205	0.1379	0.0205	1.0000
14	0.0881	460.4146	0.8746	0.4820	743.9952	0.5880	0.0371	0.5880	0.0249	0.6606
15	0.0397	209.5318	0.1488	0.6165	297.9425	0.0931	0.0162	0.0931	0.0151	0.9355
16	0.0937	348.5524	0.3352	0.6473	521.6719	0.2350	0.0344	0.2350	0.0324	0.9540
17	0.0488	588.5557	0.0488	1.0000	588.5585	0.1001	0.0183	0.1001	0.0183	1.0000
18	0.0481	233.1384	0.1574	0.6929	299.3579	0.0975	0.0224	0.0975	0.0224	1.0000
19	0.0704	1367.5564	0.2398	0.7224	2117.9765	0.1731	0.0225	0.1731	0.0225	1.0000
20	0.0713	462.3494	0.2765	0.6858	621.0968	0.1428	0.0306	0.1428	0.0306	1.0000
21	0.0704	1081.3964	0.2051	0.7484	1597.3617	0.1728	0.0232	0.1728	0.0232	1.0000
22	0.0544	2333.8907	0.1257	0.8349	2869.5270	0.1282	0.0191	0.1282	0.0191	1.0000
23	0.0881	71.3732	0.3692	0.4182	117.6898	0.2726	0.0351	0.2726	0.0219	0.6261
24	0.2489	184.6402	0.7254	0.6591	264.8534	0.5537	0.0970	0.5537	0.0955	0.9869
25	0.1320	321.0079	0.2531	0.7975	405.2692	0.2883	0.0471	0.2883	0.0471	1.0000
26	0.0464	22036.2242	0.3110	0.7547	28828.6606	0.1004	0.0178	0.1004	0.0178	1.0000
27	0.0619	8469.9885	0.0785	0.9599	8953.5581	0.1488	0.0205	0.1488	0.0205	1.0000
28	0.0447	1092.7766	0.2944	0.6104	1794.4156	0.1352	0.0163	0.1352	0.0141	0.8972
29	0.0897	1066.4611	0.1540	0.8772	1248.9623	0.1622	0.0283	0.1622	0.0283	1.0000
30	0.1175	676.3573	0.2210	0.8228	881.6454	0.2617	0.0378	0.2617	0.0378	1.0000
31	0.1672	1183.5522	0.3987	0.7918	1613.2740	0.4063	0.0542	0.4063	0.0542	1.0000
32	0.3065	20755.6167	5.8972	0.5830	95304.7125	4.9103	0.0845	4.9103	0.0178	0.6244
33	0.2818	26871.5921	0.2818	1.0000	26871.5957	0.7090	0.0776	0.7090	0.0776	1.0000
34	0.1643	1138.8055	0.6332	0.6874	1847.4319	0.3922	0.0521	0.3922	0.0521	1.0000
35	0.4780	108372.6501	3.6817	0.7742	144496.6825	1.0171	0.1613	1.0171	0.1613	1.0000
36	0.2301	621.0656	1.8679	0.5341	1057.3185	1.4274	0.0886	1.4274	0.0607	0.7021
37	0.2067	6151.3786	0.3366	0.9132	7000.0555	0.4765	0.0670	0.4765	0.0670	1.0000

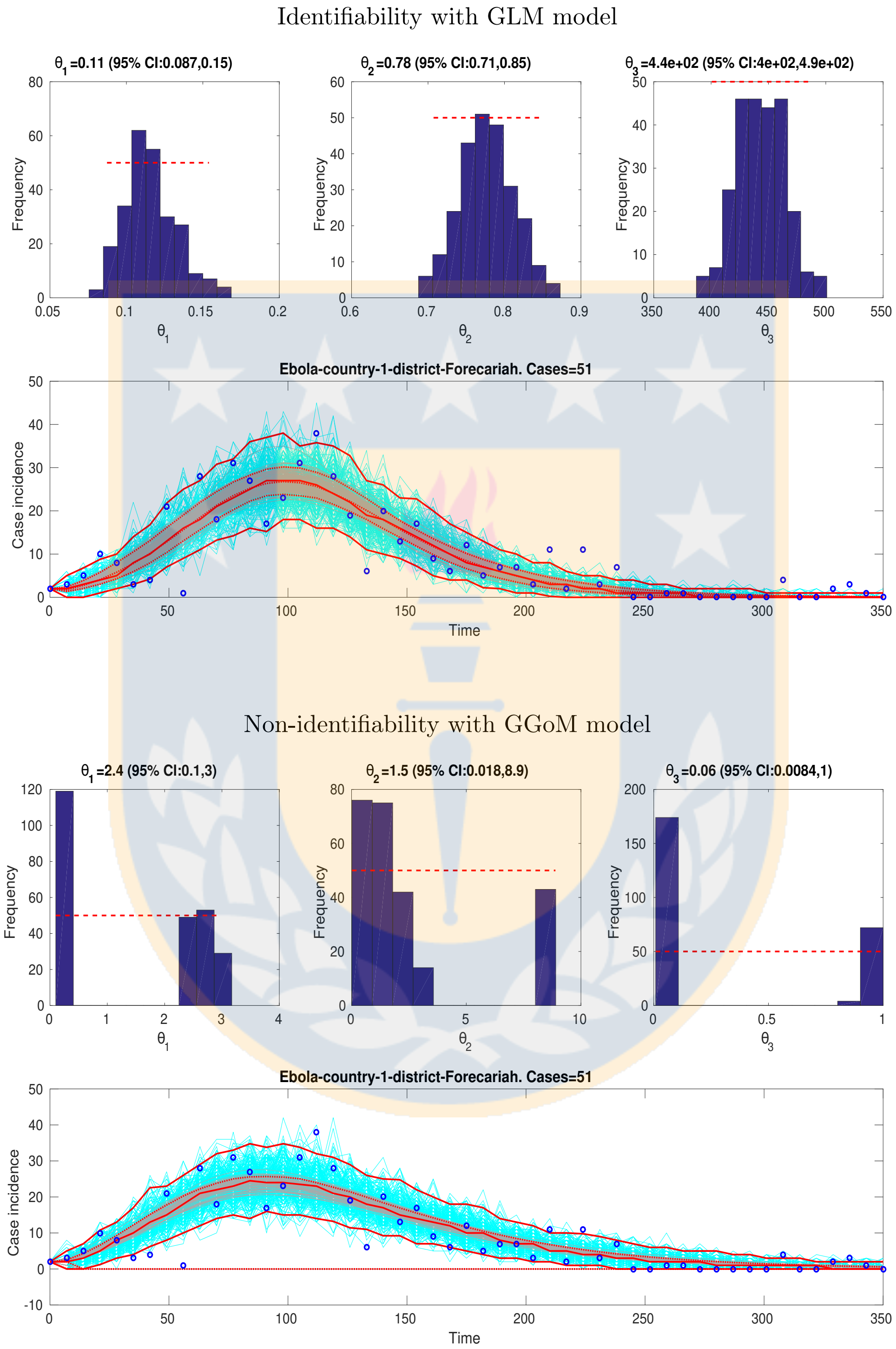
Table 1.4: Parameter estimation for LM, GLM, GoM and GGoM with total data.

Case	r		p		K	
no.	mean	95%CI	mean	95%CI	mean	95%CI
1	0.115	(0.086,0.158)	0.776	(0.696,0.846)	443.49	(394.90,487.00)
2	0.131	(0.089,0.200)	0.626	(0.526,0.717)	406.78	(365.07,452.51)
3	0.423	(0.257,0.704)	0.627	(0.471,0.777)	139.71	(119.44,163.74)
4	0.073	(0.062,0.121)	0.965	(0.790,1.000)	98.86	(82.50,116.37)
5	0.431	(0.323,0.548)	0.605	(0.559,0.671)	726.36	(668.81,778.98)
6	0.240	(0.175,0.320)	0.678	(0.606,0.773)	246.65	(218.27,274.52)
7	0.642	(0.378,1.147)	0.384	(0.227,0.523)	196.30	(166.86,220.68)
8	0.840	(0.423,2.038)	0.313	(0.005,0.499)	199.49	(160.34,283.75)
9	0.058	(0.053,0.077)	1.000	(0.885,1.000)	129.27	(108.47,150.70)
10	0.081	(0.078,0.096)	1.000	(0.950,1.000)	451.77	(417.50,494.91)
11	0.132	(0.115,0.149)	0.885	(0.853,0.922)	833.39	(775.85,894.77)
12	0.115	(0.105,0.122)	0.908	(0.894,0.928)	2556.08	(2451.96,2652.82)
13	0.152	(0.129,0.176)	0.785	(0.755,0.818)	1516.06	(1437.87,1598.21)
14	0.858	(0.638,1.192)	0.487	(0.418,0.547)	742.15	(691.23,794.99)
15	0.153	(0.098,0.216)	0.614	(0.521,0.729)	299.04	(262.92,333.23)
16	0.329	(0.252,0.443)	0.652	(0.587,0.714)	518.72	(473.90,565.42)
17	0.049	(0.047,0.059)	1.000	(0.954,1.000)	594.39	(551.33,647.65)
18	0.156	(0.100,0.235)	0.693	(0.597,0.813)	300.69	(267.36,329.84)
19	0.240	(0.216,0.265)	0.722	(0.703,0.744)	2123.24	(2028.54,2209.52)
20	0.274	(0.194,0.374)	0.688	(0.618,0.760)	620.20	(570.91,671.19)
21	0.205	(0.183,0.232)	0.749	(0.720,0.771)	1598.77	(1509.05,1682.88)
22	0.126	(0.113,0.139)	0.835	(0.815,0.855)	2869.41	(2750.68,2980.44)
23	0.356	(0.161,0.796)	0.433	(0.162,0.683)	116.16	(94.65,141.21)
24	0.719	(0.514,1.012)	0.663	(0.567,0.761)	263.51	(227.52,299.68)
25	0.256	(0.196,0.308)	0.796	(0.740,0.873)	405.35	(361.75,440.83)
26	0.310	(0.289,0.330)	0.755	(0.747,0.764)	28794.12	(28454.39,29171.09)
27	0.078	(0.074,0.083)	0.961	(0.951,0.970)	8950.33	(8758.74,9158.87)
28	0.295	(0.245,0.353)	0.609	(0.576,0.645)	1794.21	(1700.83,1870.10)
29	0.153	(0.118,0.201)	0.879	(0.819,0.938)	1250.05	(1145.82,1369.44)
30	0.221	(0.185,0.260)	0.824	(0.786,0.869)	878.98	(821.33,929.29)
31	0.400	(0.353,0.452)	0.791	(0.765,0.820)	1618.73	(1522.17,1682.63)
32	5.899	(5.227,6.851)	0.583	(0.562,0.600)	93958.96	(73179.88,139505.38)
33	3.288	(2.811,3.494)	0.639	(0.627,0.663)	26899.66	(23324.67,28492.81)
34	0.629	(0.552,0.716)	0.689	(0.665,0.714)	1848.41	(1745.74,1928.95)
35	3.700	(3.567,3.819)	0.774	(0.767,0.778)	144499.22	(88406.56,145514.69)
36	1.862	(1.483,2.435)	0.536	(0.482,0.585)	1057.80	(986.65,1116.99)
37	0.338	(0.312,0.358)	0.912	(0.902,0.927)	7016.81	(6840.94,7162.10)

Table 1.5: Confidence intervals for GLM parameters.

Case	r		b		p	
no.	mean	95%CI	mean	95%CI	mean	95%CI)
1	2.399	(0.103,2.934)	1.489	(0.017,8.886)	0.061	(0.008,1.000)
2	0.069	(0.016,0.178)	0.012	(0.010,9.034)	0.901	(0.010,1.000)
3	0.355	(0.292,0.670)	0.058	(0.043,0.074)	0.900	(0.589,1.000)
4	0.129	(0.050,0.302)	0.030	(0.020,9.357)	0.977	(0.041,1.000)
5	0.254	(0.193,0.327)	0.029	(0.026,0.033)	0.900	(0.806,1.000)
6	0.200	(0.175,0.265)	0.034	(0.028,0.040)	1.000	(0.716,1.000)
7	0.523	(0.264,1.014)	0.023	(0.018,0.032)	0.535	(0.284,0.795)
8	0.687	(0.341,1.479)	0.020	(0.011,0.032)	0.416	(0.134,0.718)
9	0.125	(0.082,1.620)	0.025	(0.018,8.948)	0.877	(0.068,1.000)
10	0.174	(0.159,0.219)	0.029	(0.026,0.032)	1.000	(0.886,1.000)
11	0.208	(0.195,0.270)	0.029	(0.027,3.012)	0.994	(0.408,1.000)
12	0.191	(0.185,0.205)	0.023	(0.022,0.024)	0.999	(0.971,1.000)
13	0.139	(0.134,0.166)	0.020	(0.019,0.021)	1.000	(0.935,1.000)
14	0.570	(0.391,0.814)	0.025	(0.022,0.029)	0.667	(0.569,0.784)
15	3.424	(0.071,4.404)	3.847	(0.014,9.774)	0.130	(0.005,0.943)
16	0.236	(0.206,0.294)	0.032	(0.029,0.036)	0.951	(0.858,1.000)
17	0.106	(0.096,0.983)	0.018	(0.015,5.936)	0.979	(0.260,1.000)
18	0.114	(0.091,3.443)	0.023	(0.018,9.747)	0.931	(0.004,1.000)
19	0.175	(0.170,0.195)	0.022	(0.021,0.023)	1.000	(0.955,1.000)
20	0.148	(0.136,0.215)	0.030	(0.026,0.032)	1.000	(0.871,1.000)
21	0.175	(0.169,0.193)	0.023	(0.021,0.024)	0.999	(0.948,1.000)
22	0.130	(0.126,0.146)	0.019	(0.018,0.019)	1.000	(0.963,1.000)
23	0.256	(0.122,0.671)	0.023	(0.014,0.035)	0.648	(0.242,1.000)
24	0.560	(0.109,3.233)	0.096	(0.077,9.604)	0.953	(0.037,1.000)
25	0.297	(0.273,0.363)	0.046	(0.040,0.050)	1.000	(0.890,1.000)
26	0.101	(0.100,0.112)	0.018	(0.017,0.018)	1.000	(0.984,1.000)
27	0.150	(0.148,0.162)	0.020	(0.020,0.021)	1.000	(0.980,1.000)
28	0.134	(0.104,0.168)	0.014	(0.013,0.016)	0.899	(0.839,0.973)
29	0.169	(0.155,0.268)	0.027	(0.019,0.030)	0.998	(0.830,1.000)
30	0.267	(0.253,0.306)	0.037	(0.034,0.039)	1.000	(0.929,1.000)
31	0.409	(0.006,3.851)	0.055	(0.050,9.469)	0.973	(0.005,1.000)
32	4.891	(3.893,6.247)	0.018	(0.011,0.025)	0.625	(0.584,0.664)
33	0.712	(0.705,0.744)	0.077	(0.076,0.078)	1.000	(0.989,1.000)
34	0.559	(0.386,3.960)	5.605	(0.049,7.869)	0.305	(0.292,1.000)
35	1.020	(1.014,1.096)	0.161	(0.159,0.162)	1.000	(0.990,1.000)
36	1.424	(1.114,1.798)	0.061	(0.054,0.069)	0.704	(0.632,0.774)
37	0.481	(0.471,0.542)	0.067	(0.065,0.068)	0.999	(0.970,1.000)

Table 1.6: Confidence intervals for GGoM parameters.



Identifiability with GLM model

Non-identifiability with GGoM model

Figure 1.9: Identifiability vs. non-identifiability of parameters for Case No 1.

CHAPTER 2

Measuring differences between phenomenological growth models applied to epidemiology

This chapter exhibits a methodology and a systematic study to measure the differences between two phenomenological growth models (PGMs). Such measures explain how such differences display the ability of certain growth models to provide a better fit to synthetic and real epidemic data.

2.1 Introduction

2.1.1 Scope

A wide variety of mathematical models have been used to study the patterns of growth processes of populations and epidemics in humans, animals, and plants [2, 13, 14, 17, 32, 61, 77, 82, 110, 127, 141, 163, 173, 179]. Here we are especially interested in dynamic growth models for characterizing epidemic trajectories, estimating key transmission parameters, gaining insight into the contribution of various transmission pathways, and providing long-term and short-term forecasts. The recent monograph by Yan and Chowell [188] provides an introduction to the topic. We herein focus on phenomenological growth models (PGMs) that only require a small number of parameters are commonly used to describe epidemic growth patterns, and which can be expressed by an ordinary differential equation (ODE) of the type

$$C'(t) := \frac{\mathrm{d}C(t)}{\mathrm{d}t} = f(t, C; \Theta), \quad t > 0; \quad C(0) = C_0, \tag{2.1}$$

where t is time, $C(t)$ is the total size of the epidemic (the cumulative number of cases) at time t, C_0 is the initial number of cases, f is an incidence function that is specific to each PGM under study, and Θ is a vector of parameters. Such models have been used to study the epidemics of influenza [5, 22, 39], Ebola [46, 47, 119, 171], Zika [19, 36, 191], Chikungunya [165], and others of global interest. The current COVID-19 pandemic is a scenario for which such models are of obvious importance [27, 55, 69, 106, 116, 135, 168, 169].

In [22] we demonstrate that some models are better at fitting data of specific epidemic outbreaks than others even when the models have the same number of parameters. Consider, for instance, the three-parameter so-called generalized logistic model (GLM) specified by

$$f(t, C; \Theta) = rC(t)^p \left(1 - \frac{C(t)}{K}\right), \quad \Theta = (r, p, K), \tag{2.2}$$

where the parameter $r > 0$ indicates the growth rate (its dimension is 1/time), K is the size of the epidemic, and $p \in [0, 1]$ is a growth scaling parameter that indicates the kind of growth (e.g., exponential vs. sub-exponential). In the comparative analysis between two models and their generalizations [22], the GLM was able to capture the trajectories for 37 real datasets describing the progression of epidemic outbreaks. In fact, this model showed to have the smallest error between the data and the fit, and the estimated parameters were identifiable, that is, the average value of each parameter was effectively a central value in the confidence intervals, where we used the definitions and calculations introduced in [32] for the error and the confidence intervals.

Although several PGMs could be considered for a given dataset, little work has been conducted to analyze the differences between models. Here we define the empirical directed distance between two PGMs as a measure of differences in the dynamics that each model is capable of generating. We address questions such as whether the dynamics of the logistic growth model (LM), defined by

$$f(t, C; \Theta) = rC(t) \left(1 - \frac{C(t)}{K}\right), \quad \Theta = (r, K), \tag{2.3}$$

is more similar to that of the Gompertz model (GoM), corresponding to

$$f(t, C; \Theta) = rC(t) \exp(-bt), \quad \Theta = (r, b), \tag{2.4}$$

where the parameter $b > 0$ describes the exponential decay of the growth rate r, or to that of the Richards model (RM)

$$f(t, C; \Theta) = rC(t) \left(1 - \left(\frac{C(t)}{K}\right)^p\right), \quad \Theta = (r, K, p). \tag{2.5}$$

We emphasize that we use the terminologies “generalized logistic model” (GLM) and “Richards model” (RM) to address *different* models, namely those given by (2.2) and (2.5), respectively. Only the model (2.5) is the one proposed originally in Richards’ paper [127]. That said, we are well aware that in parts of the literature, for instance in [63, 90, 147], the model (2.5) is referred to as “a generalized logistic model” (cf., e.g., [90]), that is “generalized logistic” and “Richards model” are used synonymously. We recall that our equation (2.2) is a generalization of the logistic model (2.3) where the exponent p is applied to the first factor $C(t)$ in (2.3), while the Richards model (2.5) represents a different generalization that arises from applying an exponent p to $C(t)/K$ within the growth-limiting factor $1 - C(t)/K$.

Before proceeding, we comment that it is arguable whether in the context of epidemiology the scalar ODE (2.1) is really a phenomenological model or simply a generator of functions to be fitted to the available data. Strictly speaking only the LM can be viewed as an epidemiological model since it arises from the well-known susceptible-infectious (SI) compartmental model in the absence of births and deaths (see, e.g., [17]). However, the use of the word 'model' for (2.1) is not only very common in the epidemiological literature (including [22,27,32,32,36,46,47,55,119,179] of the references cited so far), but we also mention that the various parameters carry relevant information characterizing the strength of an epidemic outbreak, much in contrast, say, to abstract coefficients of a function (e.g., spline function) to approximate data. That said, we emphasize that the approach of the present work is one of statistics applied to medicine and biology, and is independent of what one regards to be the 'true' status of (2.1).

There is a need to develop a methodology that helps quantify the differences in the dynamics obtained from different models that aim to capture growth processes in the social and natural sciences. Such a methodology can be helpful to assess which models are more parsimonious than others in different contexts. In the context of epidemic modeling, many models have been developed to investigate the transmission dynamics and control of infectious diseases [2, 61, 176]. However, there has not been a systematic study of differences between models and how differences in dynamics may explain the ability of certain models to provide a better fit to data than others. Here we aim to make progress in this direction by focusing on simple models that strive to capture many of the empirical patterns found in epidemic data. The main practical reason why one would be interested in understanding the distance between models stems from the need to understand how different the solutions from different models are. If two models are able to reproduce the same temporal dynamics, the researcher would be better off relying on the simpler model. Because a number of PGMs exist in the literature, we argue that understanding their differences in terms of the dynamics that they can produce adds to the literature and will help guide researchers in different applied disciplines select a reasonably small set of models rather than considering a large set of models many of which produce very similar results or fits to the data.

To address these questions we measure the differences in the dynamics between different dynamics models of the form (2.1). Here we employ simulated data for three generalized growth models (namely GLM, GGoM and RM), and with the help of mathematical and computational methods we calculate the fit and performance errors in terms of which the empirical directed distances (EDDs) are defined. As we will show, it turns out that the GLM is closer to the dynamics of the RM. On the other hand, the generalized Gompertz model (GGoM) defined by

$$f(t, C; \Theta) = rC(t)^p \exp(-bt), \quad \Theta = (r, b, p). \tag{2.6}$$

is the farthest from the RM and GLM. This is because the scaling parameter (p in (2.2), (2.5) and (2.6)) plays a more significant role in the GLM since its variation within the GLM causes more changes in its dynamics than for the other models.

The EDD between two PGMs, say A and B, is based on simulation study that we introduce in the following sections. As the foregoing discussion shows, the word "distance" in this work is not to be understood in the mathematical sense as distance function on a metric space; rather, we employ it to characterize a measure of distance based on the mean squared error. The terminology of "distance measure between models" has been employed elsewhere, cf., e.g., [159].

2.1.2 Related work

To illustrate how models can support different features of epidemic data, we can refer to the scaling of epidemic growth that characterizes the early growth dynamics of epidemics. While some epidemics spread rapidly through a population following an exponential growth phase such as pandemic influenza or the ongoing epidemic of the novel coronavirus emerging from China (COVID-19) [106], some outbreaks spread more slowly as a result of the mode of transmission or the contact network through which the pathogen spreads. For instance, sexually transmitted diseases and Ebola do not spread through the air, but require a specific type of intimate contact to spread. In such situations the disease is expected to spread follow sub-exponential growth patterns. When a model only supports exponential growth dynamics, we could expect differences between such a model and more flexible models that can capture a range of early epidemic growth dynamics [48].

The authors' interest in PGMs is mainly motivated by epidemiological applications, where the quantity that grows is usually the size of the population of infected humans. The same models also arise in quite different contexts. In fact, PGMs are commonly used in fields such as mathematical oncology and population dynamics because they consider in a simple but (up tome extent and depending on the applications) effective way phenomena concerning the growth of cells or of animal or human populations. In other words these models mirror in a simple way phenomena pertaining to these population growth phenomena. In particular they are utilized to describe the growth of a tumour where $C(t)$ is proportional to the number of cells in the tumour.

We refer to textbook entries, e.g. [65, Ch. 6], [15, Sect. 1.8], [145, p. 39], and [17, Sect. 8.2], the monograph by T.E. Wheldon [181], as well as some classic references cited in most of these works such as Aroesty et al. [6] and Newton [112]. In particular, in the latter two works it is demonstrated that the Gompertz model (2.4), under suitable choices of r and b, agrees remarkably well with data on tumour growth as long as C is not too small (as is pointed out in [65, 145]). More recent contributions that study, and compare, the applicability of various PGMs to tumour growth include [60, 62, 63, 90, 147].

One important step in our treatment consists in generating a fit of one of the PGMs to data that are either generated by another model or using real outbreak data. Since the parametric forms of PGMs are essentially non-linear, standard least-squares methods are often not applicable. Thus, to provide these fits, we resort to the Simulated Annealing (SA) method. This method is defined in [172] as a powerful stochastic search method applicable to a wide range of problems that occur in a variety of disciplines including physics, engineering problems, mathematical programming, and statistics. In particular, in the context of epidemiological models, SA has been applied to devise optimal time-profiles of public health intervention to shape voluntary vaccination for childhood diseases, see Buonomo et al. [21].

The problem can be formulated as follows. Suppose we are given finite-dimensional solution space $\mathcal{S}$, and a function $f:\mathcal{S}\to\mathbb{R}$, and we want find an optimal configuration $x^*\in\mathcal{S}$ such that $f(x^*)=\min_{x\in\mathcal{S}} f(x)$. This method has become very popular because the algorithm can solve unconstrained and bound-constrained optimization problems, especially in the multidimensional case when the objective function may have many local extremes and may not be smooth. In that case, SA is advantageous because it does not require calculation of derivatives, and thus be considered as a derivative-free method. In papers including [87, 128] this method has been used for parameter estimation, which motivated our computation.

2.2 Distance between phenomenological growth models

2.2.1 Notation and solution of PGMs

We begin the discussion with a comment on notation. The notation chosen for the incidence function $f=f(t,C;\Theta)$ presupposes that t and C are independent arguments. In fact, it is also possible to rewrite all models utilized herein as autonomous ordinary differential equations of the form

$$C'(t)=\varphi(C;\Theta),\quad t>0;\quad C(0)=C_0. \tag{2.7}$$

This is directly obvious for the GLM (2.2) and the RM (2.5). Rewriting the ODEs for the GoM and GGoM (with $0<p<1$) as

$$\exp(-b\tau)=\frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}\tau}\ln C(\tau)\quad\text{and}\quad \exp(-b\tau)=\frac{1}{r(1-p)}\frac{\mathrm{d}}{\mathrm{d}\tau}C^{1-p}\quad\text{for } 0\leq\tau\leq t,$$

integrating with respect to τ, and substituting the corresponding expression for $\exp(-bt)$ in (2.4) and (2.6), respectively, one obtains

$$\varphi(C;\Theta)=\bigl(r+b\ln C_0\bigr)C-bC\ln C$$

for the GoM and

$$\varphi(C;\Theta)=C^p\left(\frac{b}{p-1}\bigl(C^{1-p}-C_0^{1-p}\bigr)+r\right),\quad 0<p<1,$$

for the GGoM. The original and autonomous forms, (2.1) and (2.7), are of course equivalent in all cases. However, one or another form is preferable depending on the context of application of

Phenomenological growth model	Parameters
Logistic growth model (LM)	$\Theta = (\vartheta_1 = r, \vartheta_2 = K);\ r, K > 0$
Generalized Logistic growth model (GLM)	$\Theta = (\vartheta_1 = r, \vartheta_2 = K, \vartheta_3 = p);\ r, K > 0,\ p \in [0,1]$
Richards model (RM)	$\Theta = (\vartheta_1 = r, \vartheta_2 = K, \vartheta_3 = p);\ r, K > 0,\ p \in [0,1]$
Generalized Gompertz model (GGoM)	$\Theta = \{\vartheta_1 = r, \vartheta_2 = b, \vartheta_3 = p\};\ r, b > 0,\ p \in [0,1]$

Table 2.1: Summary of information on models and parameters.

the respective PGM, that is, on whether dependence of the growth rate on time or on the current size of population should be emphasized. For instance, as is pointed out in [6], in the application to tumour growth it considered more suggestive to relate the specific growth rate for a particular tumour to its size. On the other hand, in the application to epidemiological data we wish to compare model curves to given time series of data. Furthermore, we prefer a formulation that allows one to include explicit time dependence of the incidence rate by external factors at a later stage. One such external factor could be, for example, the seasonal variation of temperature. As compromise between these different viewpoints, we have chosen the dependence $f = f(t, C; \Theta)$. Finally, we mention that additional insight and comparison between different models can also be achieved from considering the population size and its growth rate as separate state variables and analyzing the fixed points and stability of the resulting dynamical system of two scalar equations, as is done e.g. in [60]. For instance, the GGoM (2.6) can be written as the coupled system $C' = aC^p$, $a' = -ba$ with the initial conditions $C(0) = C_0$, $a(0) = r$.

For the growth models summarized in Table 2.1, $p = 0$ corresponds to a constant incidence over time, $p = 1$ corresponds to exponential growth, and any intermediate value $0 < p < 1$ leads to a model that describes initial sub-exponential growth dynamics [42, 45, 48, 170]. In fact, in prior work we have demonstrated that some epidemics are characterized by early slower-than-exponential growth using flexible phenomenological models (see [45, 170]).

Three of these models have an initial logistic growth because when $p = 1$ for the GLM and RM, in other words the LM is recovered. In contrast, this is not the case for the GGoM. (The RM and GLM show two forms of incorporating the parameter p to the LM model to obtain the generalized growth form $rC^p(t)$.) We wish to measure how close the logistic models are to each other and to the GGoM, and to assess whether two or three parameters are sufficient to recover other dynamics. We recall the following explicit solutions. The solution of the LM (2.1), (2.3) is given by

$$C(t) = \frac{KC(0)\exp(rt)}{K + C(0)(\exp(rt) - 1)}, \tag{2.8}$$

that of the GoM (2.1), (2.4) (that is, the GGoM for $p = 1$) by

$$C(t) = C(0)\exp\bigl((r/b)\bigl(1 - \exp(-bt)\bigr)\bigr), \tag{2.9}$$

while for the GGoM (2.1), (2.4) we get

$$C(t) = \big((1-p)(r/b)\big(1-\exp(-bt)\big) + C(0)^{1-p}\big)^{1/(1-p)} \quad (\text{where } 0<p<1). \tag{2.10}$$

The solution of the RM (2.1), (2.5) is

$$C(t) = \frac{KC(0)\exp(rt)}{(K^p + C(0)^p(\exp(prt)-1))^{1/p}}. \tag{2.11}$$

As is pointed out in [22], the GLM (2.1), (2.2) does not have a solution in closed algebraic form for general values $p \in (0,1)$. (This point is also discussed in detail in [117]; the Pütter-Bertalanffy growth equation studied in that paper includes (2.1), (2.2) as a special case.) For the GLM we solve the initial-value problem (2.1) numerically whenever necessary.

Phenomenological growth models can capture epidemic growth patterns, through the relationship between the case incidence curve and the cumulative incidence curve. The integrated version of (2.1), namely

$$C(t) = C(0) + \int_0^t f(\tau; C; \Theta)\,\mathrm{d}\tau, \quad t > 0,$$

can be approximated by the following formula if we assume that values of the incidence function $f(t, C; \Theta)$ are given at discrete times $t = t_k$, $k = 1, \dots, n$ only:

$$C(t_k) \approx C(0) + \sum_{l=1}^{k} (t_l - t_{l-1}) f(t_l, C; \Theta), \quad k = 1, 2, \dots, n, \quad t_0 = 0,$$

with $t_k \in [0, T]$. Thus, we may recover the cumulative curve $t \mapsto C(t)$ in terms of tabulated values of the incidence function $f(t, C; \Theta)$, and similarly we may approximate $f(t_k; C; \Theta)$ in terms of given discrete values $C(t_k)$ as follows:

$$f(t_k; \Theta) \approx \frac{C(t_k) - C(t_{k-1})}{t_k - t_{k-1}}, \qquad k = 1, 2, \dots, n, \qquad \text{with} \qquad f(t_0; C; \Theta) = C(t_0). \tag{2.12}$$

2.2.2 Measuring the distance between PGMs

To determine EDD($B \to A$), we start by defining S parameter sets Θ_j, $j = 1, \dots, S$ for model A for which we determine the incidence curves, that is, we compute the (exact or numerical) solutions for the ODE (2.1) for model A for each parameter set Θ_j, and these are our datasets to fit model B. We fit model B to each of these curves by using Q different initial parameter sets to execute a numerical program that applies a process of minimization to estimate parameters of model B. These initial parameter sets, in turn, are created by using a method of latin hypercube sampling that creates Q random values within a defined range. For instance, for the parameter K we create $Q = 10$ values between 0 and 1000. Assume now that y_{t_i}, $i = 1, \dots, n$, are the points or data for each time t_i of model A, and $f(t_i, C; \hat{\Theta})$, are the

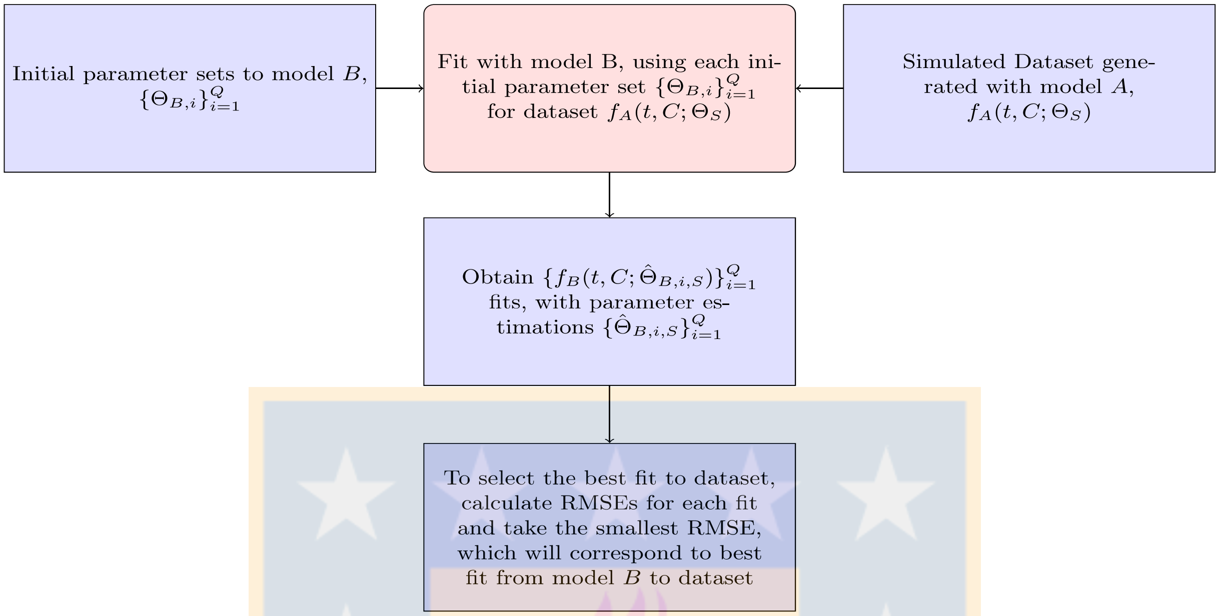

Initial parameter sets to model B, $\{\Theta_{B,i}\}_{i=1}^{Q}$
Fit with model B, using each initial parameter set $\{\Theta_{B,i}\}_{i=1}^{Q}$ for dataset $f_A(t,C;\Theta_S)$
Simulated Dataset generated with model A, $f_A(t,C;\Theta_S)$
Obtain $\{f_B(t,C;\hat{\Theta}_{B,i,S})\}_{i=1}^{Q}$ fits, with parameter estimations $\{\hat{\Theta}_{B,i,S}\}_{i=1}^{Q}$
To select the best fit to dataset, calculate RMSEs for each fit and take the smallest RMSE, which will correspond to best fit from model B to dataset

Figure 2.1: Process to fit the model B to dataset $f_A(t,C;\Theta_S)$ generated with model A.

values of fits obtained with model B, where $\hat{\Theta}$ is the set of estimated parameters of model B. Then we determine the root mean square error (RMSE)

$$\text{RMSE} := \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(f(t_i,C;\hat{\Theta}) - y_{t_i}\right)^2}$$

to compute the distance between the data curve, tabulated at $t_1,\ldots,t_n$, and a fit with model B expressed by the values $f(t_i,C;\hat{\Theta})$, $i=1,\ldots,n$. We select the best fit with the smallest RMSE between the Q fits for each of the S data curves (see Figure 2.1), and then consider the mean of the S best values of RSME as the distance from model B to model A. Besides, we will also calculate the sum of squared errors (SSE) given by

$$\text{SSE} = \sum_{i=1}^{n}\left(f(t_i,C;\hat{\Theta}) - y_{t_i}\right)^2,$$

because this quantity naturally arises in the context of least-squares methods. The necessary computations are summarized in Algorithm 1 and in Figure 2.2.

2.2.3 Simulated Annealing method for parameter estimation

As we want know the distance between two PGMs, we need numerical methods to calculate the fits from a model B to a model A and in some cases to determine solutions of the ODEs. Then to achieve the best fit it is necessary to estimate parameters, for which we employ

Algorithm 1: Calculating EDD$(B \to A)$

```
Input:
  • Parameter sets $\{\Theta_j\}_{j=1,\dots,S}$ of model $A$ that define the incidence curves
        $\{f_A(t, C; \Theta_j)\}_{j=1,\dots,S}, \quad t \in [0, T]$
    (simulated data).
  • Initial parameter sets $\{\Theta_{B,i}\}_{i=1,\dots,Q}$ of model $B$.
  • Sampling times $t_k$, $k = 1, \dots, n$ at which the incidence curves (both of the simulated
    data and the approximation) are evaluated.
for $j = 1$ to $S$ do
  $i^*(j) \leftarrow 1$
  for $i = 1, \dots, Q$ do
(1)       Determine the vector of estimated parameters $\hat{\Theta}_{B,i,j}$ for the $j$-th data curve
          based on the initial parameter vector $\Theta_{B,i}$ by SimulatedAnnealing.
(2)       Calculate
          $\mathrm{RMSE}_{ij} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\bigl(f_B(t_k; C; \hat{\Theta}_{B,i,j}) - f_A(t_k; C; \Theta_j)\bigr)^2}.$
    if $\mathrm{RMSE}_{ij} \leq \mathrm{RMSE}_{i^*(j),j}$ then
          $i^*(j) \leftarrow i$
    end if
  end for
end for
Output: the empirical directed distance from model $B$ to model $A$,
          $\mathrm{EDD}(B \to A) \leftarrow \frac{1}{S}\sum_{j=1}^{S} \mathrm{RSME}_{i^*(j),j}.$
```

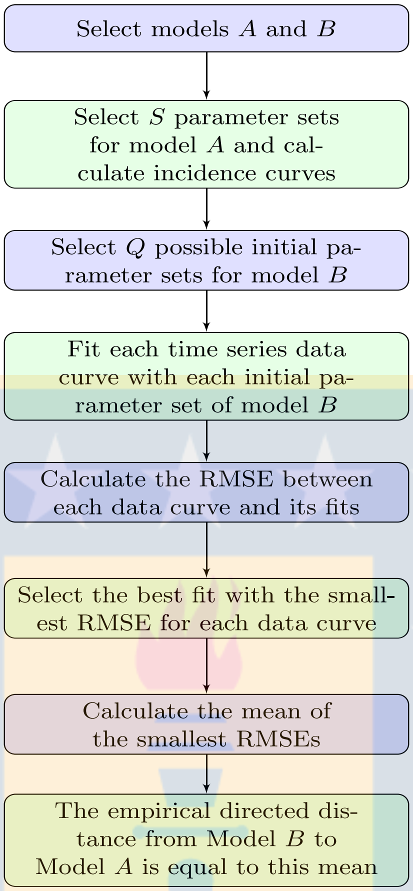

Select models A and B
Select S parameter sets for model A and calculate incidence curves
Select Q possible initial parameter sets for model B
Fit each time series data curve with each initial parameter set of model B
Calculate the RMSE between each data curve and its fits
Select the best fit with the smallest RMSE for each data curve
Calculate the mean of the smallest RMSEs
The empirical directed distance from Model B to Model A is equal to this mean

Figure 2.2: Step by step of measuring the empirical directed distance between two models.

the Simulated Annealing method to minimize the Euclidean distance between the curve from model A and the fit with the estimation parameters of model B. The Simulated Annealing (SA) method has been useful to solve optimization problems [18], in particular for parameter estimation [1, 50, 128], as in our case, where the goal here is to minimize the function

$$\Theta \mapsto J(\Theta) := \sqrt{\sum_{k=1}^{n} \big(f(t_k; C; \Theta) - \mathrm{data}_{t_k} \big)^2},$$

being $t \mapsto f(t, C; \Theta)$ the incidence function of a PGM evaluated for a parameter vector Θ that should satisfy $\Theta \in \mathcal{S}$ for some admissible set $\mathcal{S}$ compatible with the algebraic form of f for n different time points t_k, where data_{t_k} correspond to data in time series. In our study, the values $\{\mathrm{data}_{t_k}\}_{k=1,\ldots,n}$ are the datasets generated by model A, and model B will define the incidence function f and the set $\mathcal{S}$. Hence, the optimization problem at hand can be defined as follows:

$$\text{find } \hat{\Theta} \in \mathcal{S} \text{ such that } J(\hat{\Theta}) = \min_{\Theta \in \mathcal{S}} J(\Theta). \tag{2.13}$$

This problem is solved by employing the routines exposed in Appendix A.

2.3 Application of the methodology

2.3.1 Parameters of specific phenomenological growth models

The methodology of Section 2.2.2 will allow us to determine the contribution of the scaling parameter p and to observe the closeness between the dynamics of the models A and B, where model $A \in \{\mathrm{GLM}, \mathrm{RM}, \mathrm{GGoM}\}$ is used to generate simulated data or data curves and model $B \in \{\mathrm{LM}, \mathrm{GLM}, \mathrm{RM}, \mathrm{GGoM}\}$, $B \neq A$, is employed to calculate fits. To assess the contribution of the parameter p, we select a set of values of p fairly close to 1 but leave other parameters fixed (taking into account that the parameter b of the GGoM depends on the value of p). Then, we analyze the distance of model B to curves generated with model A. For example, if we consider $B = \mathrm{LM}$ and its fits to each data curve generated with model A, we can calculate the RMSEs, and finally to have the distance from the LM to GLM, RM and GGoM curves. Furthermore, we also calculate the distance from the GLM to RM and GGoM curves, the RM to GLM and GGoM curves and finally from the GGoM m to GLM and RM curves. These processes will be named Experiment 1, 2, 3, and 4, respectively. (All these distances are to be understood in the sense of EDD, of course.)

For the experiments we consider the three parameters r, p, and K. To compare models with equivalent parameters, we choose (as in [22, Sect. 1]) the following expressions for the parameters b and r within the GGoM in terms of the parameter K and the initial value $C(0)$:

$$r = 1 - \frac{C(0)}{K}, \tag{2.14}$$

$$b = \begin{cases} \dfrac{r}{\log(K/C(0))} & \text{if } p = 1, \\ \dfrac{r(1-p)}{K^{1-p} - C(0)^{1-p}} & \text{if } 0 < p < 1, \end{cases} \tag{2.15}$$

where the expression for $p = 1$ is the limit of that for $0 < p < 1$, i.e.,

$$\frac{r}{\log(K/C(0))} = \lim_{p \to 1, p<1} \frac{r(1-p)}{K^{1-p} - C(0)^{1-p}}.$$

Therefore, to standardize the analysis, we consider the parameter set $\Theta = (r, p, K)$ for all models with $K = 1000$, $C(0) = 1$, r determined by (2.14), and b calculated from (2.15) in dependence of the value of the parameter p, which is allowed to assume one of the values

$$p \in \mathcal{P} := \{1, 0.995, 0.99, 0.98, 0.95, 0.85, 0.8\}.$$

Summarizing, we utilize the parameters

$$(r, p, K) = (0.999, p, 1000) \quad \text{with } p \in \mathcal{P}.$$

Parameters for GGoM curves			Parameters for RM curves			Parameters for GLM curves		
r	b	p	r	p	K	r	p	K
0.999	0.1446	1.000	0.999	1.000	1000	0.999	1.000	1000
0.999	0.1421	0.995	0.999	0.995	1000	0.999	0.995	1000
0.999	0.1397	0.990	0.999	0.990	1000	0.999	0.990	1000
0.999	0.1349	0.980	0.999	0.980	1000	0.999	0.980	1000
0.999	0.1211	0.950	0.999	0.950	1000	0.999	0.950	1000
0.999	0.0824	0.850	0.999	0.850	1000	0.999	0.850	1000
0.999	0.0670	0.800	0.999	0.800	1000	0.999	0.800	1000

Table 2.2: Summary of parameters for each model curve

These values are used directly for the GLM and RM, while for the GGoM we employ the corresponding parameters $(r, p, b) = (0.999, p, b)$ with $b = 0.1446$ if $p = 1$ and $b = 0.1421$, 0.1397, 0.1349, 0.1211, 0.0824, and 0.0670 for $p = 0.995$, 0.99, 0.98, 0.95, 0.85, and 0.8, respectively.

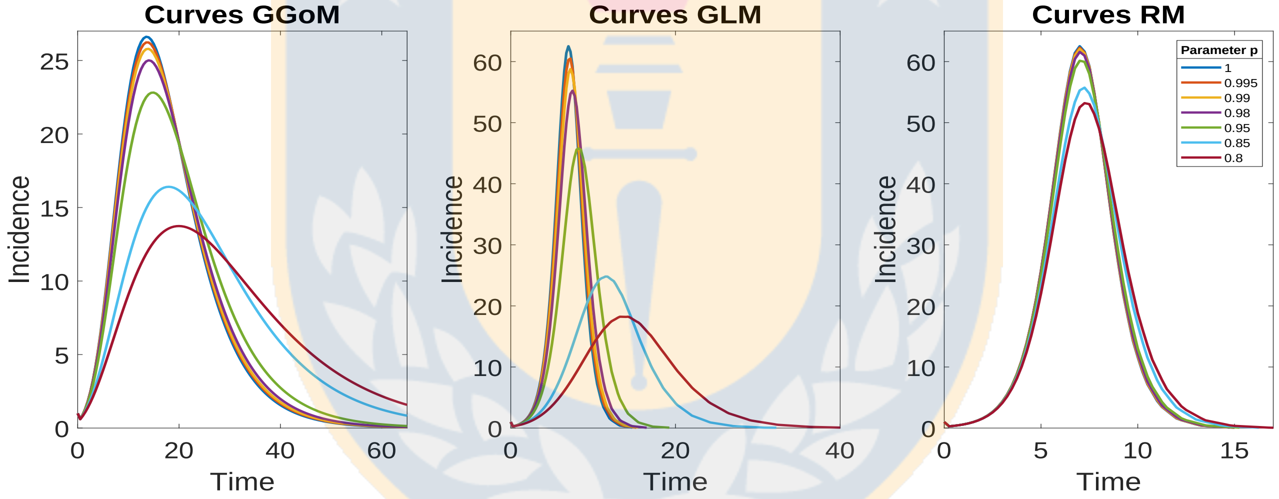


Figure 2.3: Data curves for each growth model.

These parameters, listed also in Table 2.2, produce the data curves shown in Figure 2.3. Roughly speaking, these curves illustrate that the role of the parameter p within the GGoM and GLM is to describe the initial growth of the incidence curve, while within the RM the initial phase of the curves, where values of $C(t)$ are still fairly small, is almost the same for all p-values. Furthermore, we see that with decreasing values of p the extremal value (peak) of the incidence curves of the GGoM and GLM decreases rapidly, while that of the RM model decreases only slowly. In addition, the GGoM and GLM exhibit an appreciable shift of the timing of that maximum (i.e., the peak time increases significantly with decreasing p) while this effect is not much appreciable for the RM (with the chosen parameters). (For the GGoM and RM the respective closed formulas for $C(t)$, (2.10) and (2.11), may be utilized and differentiated to

Model B			
LM $\Theta = (r, K)$	GLM $\Theta = (r, p, K)$	RM $\Theta = (r, p, K)$	GGoM $\Theta = (r, b, p)$
$(0, 4) \times (0, 1000]$	$(0, 4) \times (0, 1] \times (0, 1000]$	$(0, 4) \times (0, 1] \times (0, 1000]$	$(0, 4) \times (0, 1] \times (0, 1]$

Table 2.3: Initial parameter set for each model B

Model A / Model B	GLM curves	RM curves	GGoM curves
LM (r, K)	$[0.5, 1.1] \times (700, 1010)$	$[0.9, 1.1] \times (900, 1010)$	$[0.2, 0.5] \times [400, 800]$
GLM (r, p, K)		$[0.8, 1] \times [0.2, 1] \times [500, 1010)$, $0.99 \leq p \leq 1$; $[0.5, 1.5] \times [0.4, 0.85] \times [800, 1010)$, $0.95 \leq p < 0.99$; $[0.5, 1] \times [0.4, 0.999] \times [900, 1010)$, $0.8 \leq p < 0.95$	$[1.5, 1.6] \times [0.5, 0.7] \times [800, 1000]$
RM (r, p, K)	$[0.7, 0.99] \times [0.2, 0.999] \times [800, 1010)$		$[1.8, 1.9] \times [0.05, 0.08] \times [800, 1000]$, $0.95 \leq p \leq 1$ $[1.8, 2] \times [0.05, 0.08] \times [800, 1010)$, $0.8 \leq p \leq 0.25$
GGoM (r, b, p)	$(0, 3) \times (0, 1] \times (0, 1]$	$(0, 3) \times (0, 1] \times (0, 1]$	

Table 2.4: Solution spaces for the parameter estimation with each model B and data curves.

discuss all these properties in explicit form, see [22].)

To help the fits, we generate the data curves from model A, with evaluations for every $0 < h < 1$ time units to have more points or data for fit model B in each case, i.e. we select $t_k = kh$, $k = 0, 1, 2, 3, \ldots, n$. For example, we use temporal meshwidth of $h = 0.25$ for the GLM curves.

2.3.2 Application of the Simulated Annealing (SA) method

The SA method will be used to estimate parameters, as presented in appendix A, where we will use the Matlab function SIMULANNEALBND to implement the SA algorithm. The objective function to minimize is (2.13) for parameter vectors Θ and functions f that depend on the choice of model B in each case. For simplicity, the application of SA method, we will use the solutions from model B, where by utilizing (2.12), we could recover f in terms of C.

For example, the function f within the objective function for $B = \text{LM}$ is calculated by using the solution C to the LM presented in (2.8), i.e. we use the explicit solution of this model, as we also do for the RM with (2.11) and the GGoM with (2.9) and (2.10) for the respective cases $p = 1$ and $0 < p < 1$. However, since the GLM does not have a solution in closed algebraic form we employ a numerical approximation to solve the initial value problem to the GLM, as is detailed in Appendix A.

Then, once the form of the algebraic model under study is given, we need to define the solution spaces for each model which depend on the role of each parameter within each model function.

		Error RMSE to each fit with model B							
Model B	Model A	$p = 1$	$p = 0.995$	$p = 0.99$	$p = 0.98$	$p = 0.95$	$p = 0.85$	$p = 0.8$	EDD$(B \to A)$
LM	GGoM	5.2319	5.1864	5.1430	5.1163	5.0480	4.7069	4.4697	5.1163
	GLM	0.1900	0.2455	0.4625	0.8184	1.7021	2.6900	2.6570	0.8184
	RM	0.0568	0.0685	0.0955	0.1706	0.4099	1.1989	1.5615	0.1706
RM	GGoM	0.6827	0.6804	0.7055	0.7668	0.9285	1.3375	1.4244	0.7668
	GLM	0.0037	0.0381	0.0741	0.1347	0.2638	0.3397	0.3066	0.1347
GLM	GGoM	0.4712	0.4757	0.4556	0.4481	0.4284	0.3513	0.3015	0.4481
	RM	0.0069	0.1536	0.0605	0.1993	0.2477	0.7268	1.6235	0.1993
GGoM	GLM	12.1578	11.6221	11.1988	10.2038	7.8028	3.4788	1.8623	10.2038
	RM	12.1667	12.1529	12.0017	12.0359	11.4630	10.1060	9.3656	12.0017

Table 2.5: RSME for each data curve, where columns 3 to 9 correspond to the error for the indicated value of p, and column 10 shows the mean RMSE, that is, EDD$(B \to A)$.

Here the quantities K, $C(0)$, and p are fixed and the expressions for the parameters r and b are given by (2.14) and (2.15), respectively. To search the solution spaces for each parameter, we consider the conditions summarized in Table 2.1 to define the sets specified in Table 2.3, where we select the initial parameter to run the SA algorithm. This algorithm provides a solution that varies from run to run since the algorithm consists in a random process that utilizes a probability criterion to select the optimal value. However, if we apply the SA algorithm to Q possible initial parameter sets, then with these solutions we can reduce or limit the solution space between the maximum and the minimum best parameters shown for the run. This new solution space helps us to control results and improve the solution and the calculation time. This process follows the idea shown in [128] concerning double-cycle application of SA. The solution spaces that result from the fits for each model B with each data curve are summarized in Table 2.4.

2.3.3 Experiment 1: empirical directed distances from the logistic model (LM) to other models

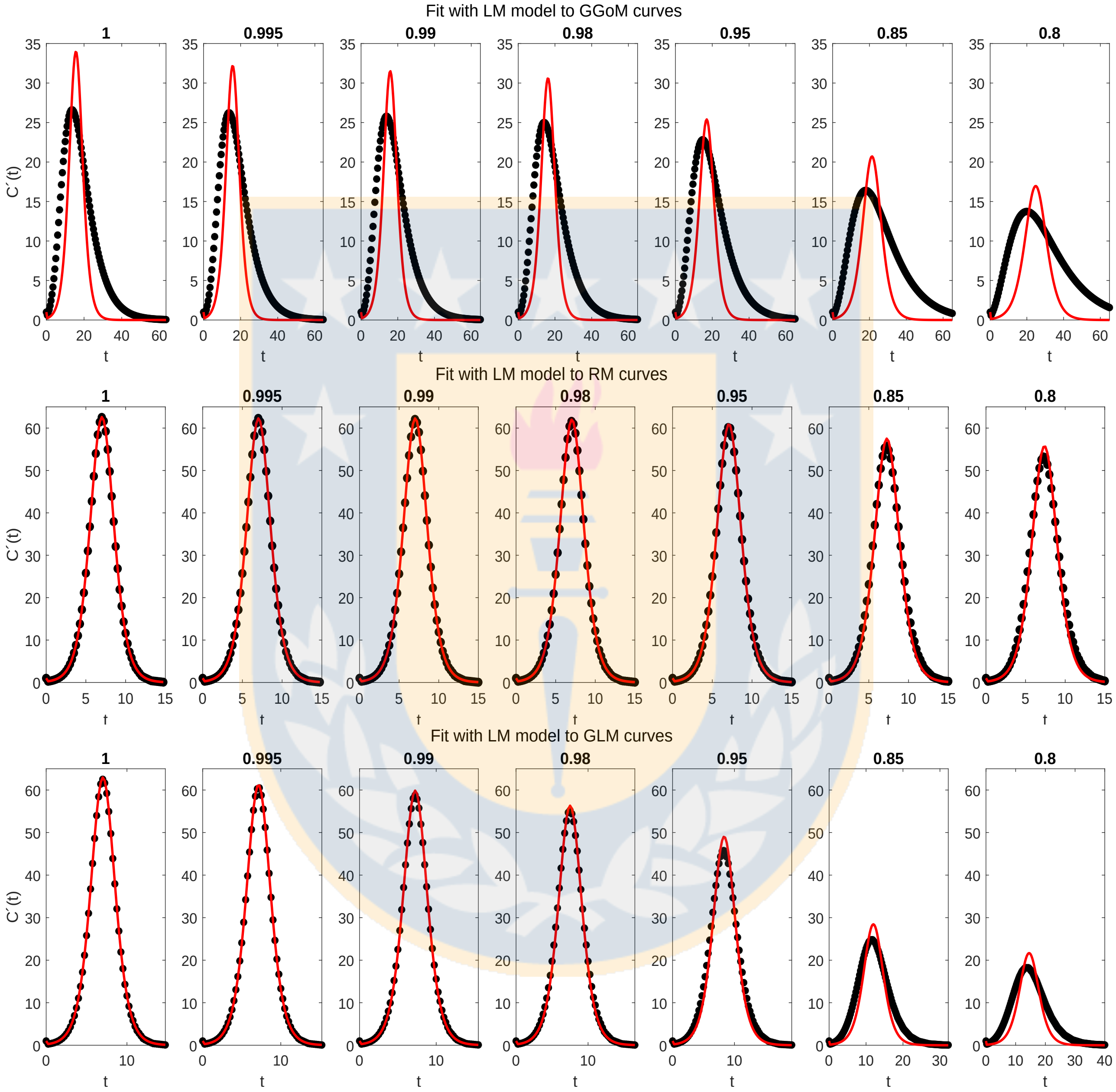


Figure 2.4: Experiment 1: results of fits of the LM (model B) to the curves of data generated by the GGoM (top row), GLM (middle row), and GLM (bottom row), for the indicated values of p.

Parameter Estimation for LM						
CURVES	with GGoM curves		with GLM curves		with RM curves	
with p	r	K	r	K	r	K
1	0.4193	621.2245	1.0017	1007.1734	1.0012	1001.6787
0.995	0.4149	617.9446	0.9836	997.3534	0.9999	998.2173
0.99	0.4105	617.1784	0.9662	992.6324	0.9989	999.1522
0.98	0.4018	610.0832	0.9310	969.1496	0.9965	996.8109
0.95	0.3756	598.7651	0.8341	941.6386	0.9888	987.6975
0.85	0.2931	554.6741	0.5643	806.5594	0.9600	959.6222
0.8	0.2550	535.9005	0.4597	754.1879	0.9433	946.7694

Table 2.6: Experiment 1: parameter estimation for LM with GGoM, GLM and RM data curves.

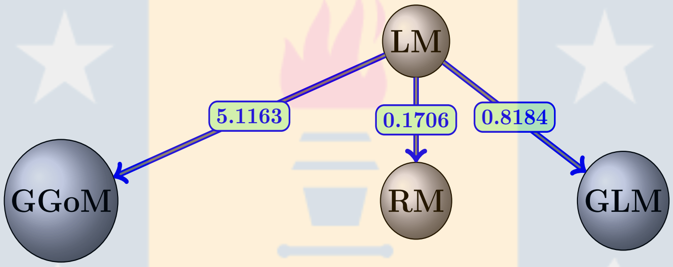


Figure 2.5: Experiment 1: illustrative diagram for the empirical directed distances EDD(LM → GGoM), EDD(LM → RM), and EDD(LM → GLM) based on data curves.

With the best set of initial parameters and the best parameter estimation, we have Figure 2.4 with the best fits for the LM, where we can see that the LM is closer to the RM curves, since it captures this dynamics better than for that of the other models. On the other hand, LM is further from GGoM curves, this is due to the long time defined for GGoM data, that the LM exceeds the maximum given by it. A similar situation occurs when the maximum decreases for GLM curves and time increases. The RMSEs calculated to measure the EDD are shown in Table 2.5 and Figure 2.5. It turns out that that when the value of p is decreased for the GLM and RM, the error increases more for the GLM than for the RM while a different situation occurs with the GGoM, since the error decreases when p is decreased, but this change is slower than the increase of the error for the GLM and RM. The increase of the RMSE for data generated by the LM is expected because when $p = 1$, the dynamics of the LM and that of these models should be the same, where in Table 2.6 (first row) we can see that the parameter estimation for GLM and RM data curves with $p = 1$ are closer to real parameters, i.e., to $\Theta = (r = 0.999, p = 1, K = 1000)$. Another observation about results for parameter estimation summarized in Table 2.6 is that the growth rate r of the LM for data curves generated by the GGoM is naturally smaller than the growth rate for data generated by the LM, because the

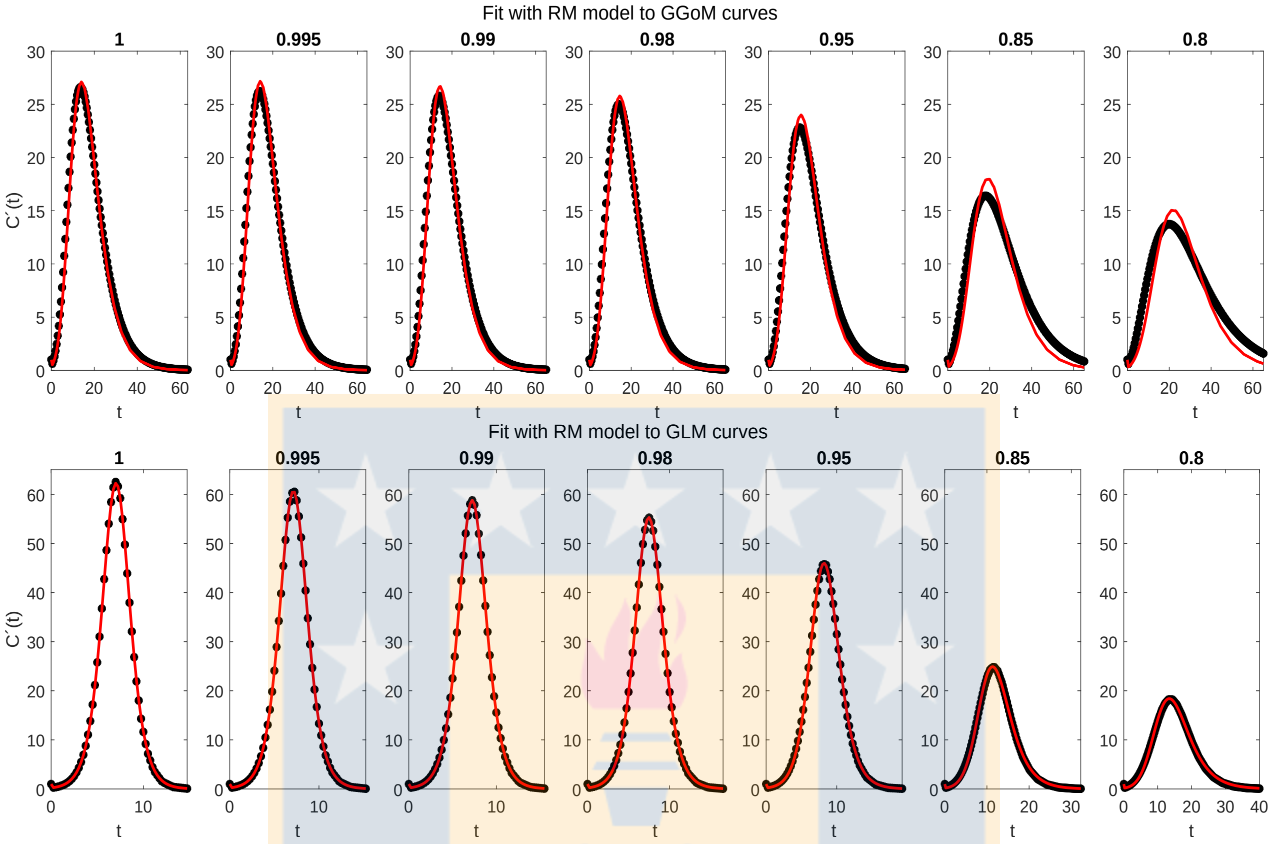


Figure 2.6: Experiment 2: results of fits of the RM (model B) to the curves of data generated by the GGoM (top row) and the GLM (bottom row), for the indicated values of p.

GGoM has a slower increase, where for the same reason for GLM data with $p = 0.8$ the growth rate decreases to 0.4597.

2.3.4 Experiment 2: empirical directed distances from the Richards model (RM) to other models

We follow the structure of presentation of results of Experiment 1. In Figure 2.6 we can observe that the RM (in the role of model B) is closer to the GLM than to the GGoM, where the fits captures almost all the dynamics presented for the GLM data curves. Now with the RMSE calculated, we have effectively the smallest errors for the fit to GLM data, where in Table 2.5 we see that the RMSE increases faster with GGoM data than with GLM data. Besides, the RMSEs for GLM curves are less than 0.5, evidencing relative closeness between the logistic models. Concerning the parameter estimation (Table 2.7), we have a good approximation between the parameters for GLM when $p = 1$, where the estimated parameter p varies more than the growth rate r to capture the decrease of the maximum value, evidencing a good contribution of this parameter. On the other hand the variation of the parameter r is smaller than that of p and K when the RM is used to fit the GGoM curves.

Parameter Estimation for RM						
CURVES	with GGoM curves			with GLM curves		
with p	r	p	K	r	p	K
1	1.9998	0.0800	954.1139	0.9990	0.9999	1000.0513
0.995	1.9999	0.0798	957.9978	0.9876	0.9732	999.5539
0.99	1.9973	0.0789	953.8977	0.9763	0.9476	999.4866
0.98	1.9912	0.0771	944.5345	0.9547	0.8976	997.8995
0.95	1.9431	0.0740	937.1508	0.9000	0.7532	1000.3445
0.85	1.8952	0.0593	892.6791	0.8000	0.4048	998.1613
0.8	1.9238	0.0500	869.3741	0.8551	0.2612	1003.6135

Table 2.7: Experiment 2: parameter estimation for RM with GGoM and GLM data curves.

2.3.5 Experiment 3: empirical directed distances from the generalized logistic model (GLM) to other models

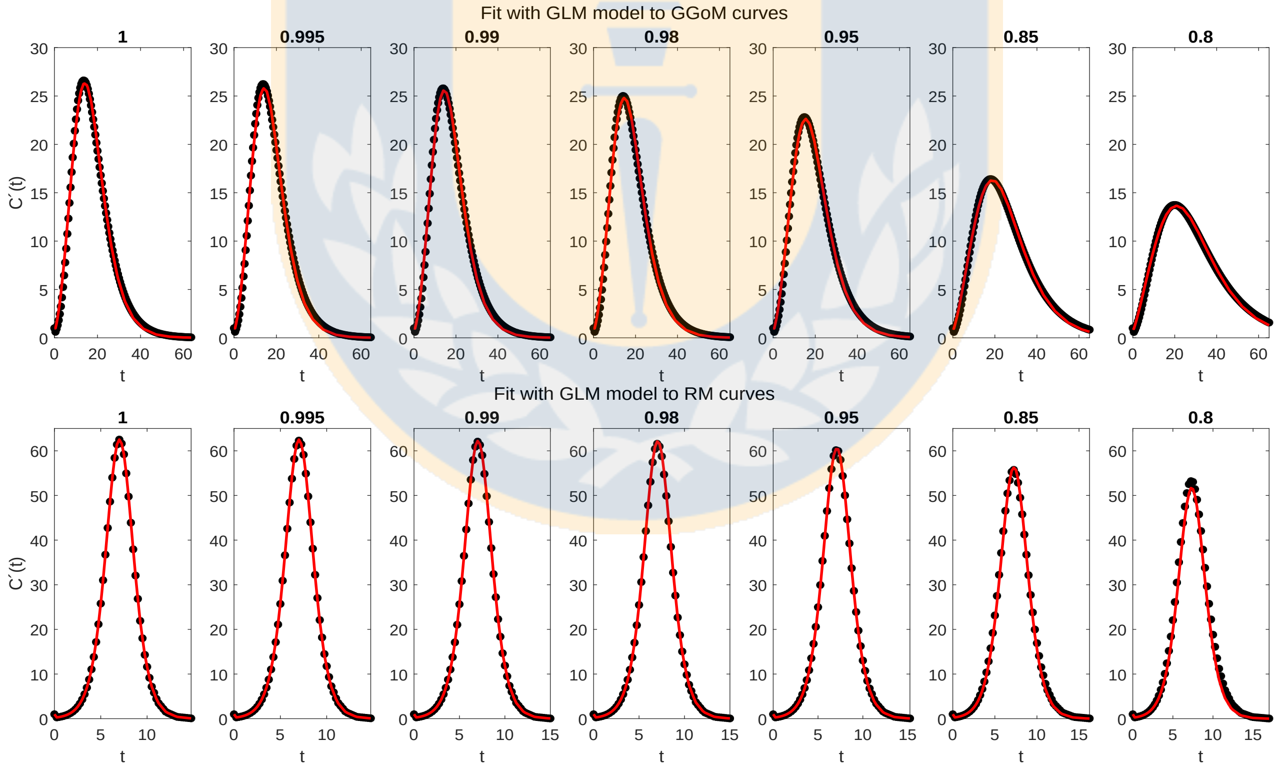


Figure 2.7: Experiment 3: results of fits of the GLM (model B) to the curves of data generated by the GGoM (top row) and the RM (bottom row), for the indicated values of p.

Parameter Estimation for GLM						
CURVES	with GGoM curves			with RM curves		
with p	r	p	K	r	p	K
1	1.5402	0.6744	998.0430	0.9994	0.9999	1000.1965
0.995	1.5293	0.6734	986.8504	1.0000	0.9999	999.3734
0.99	1.5054	0.6743	989.5241	1.0000	0.9991	1000.4114
0.98	1.5221	0.6670	992.3809	1.0000	0.9984	1003.5106
0.95	1.5278	0.6508	992.7939	1.0000	0.9961	992.1099
0.85	1.5204	0.5962	994.5254	1.0000	0.9876	971.7100
0.8	1.5112	0.5678	989.0751	0.9888	0.9850	915.3278

Table 2.8: Experiment 3: parameter estimation for GLM with GGoM and RM data curves.

In Figure 2.7, we can see a performance closer to both dynamics with GLM, where this model captures fairly well the maximum value and the length time. Observing the RMSEs (2.5), we can see that these are smalller than 1.6, as expected when we consider the fits shown in Figure 2.7. Now, analyzing Table 2.5 we observe that the errors increase faster for RM (when p decreases) than with GGoM, where the errors decrease slowly when p decreases. This behavior may be due to the dynamics of the GLM, where if the maximum value decreases, the time length increases, but for the RM data curves, the time length and maximum value are closer to each other. About parameter estimation (see Table 2.8), we have that for the parameter set with the RM curves the values are closer to parameters of the GLM with $p = 1$, i.e, $\Theta = (0.999, 1, 1000)$. This, because, the RM curves vary little of the RM initial curve with $p = 1$. The previous result contrasts with the fit for GGoM curves, because when the parameter p varies for GGoM curves, the maximum value decreases and the time length increases, where with the GGoM the length time is the same when the parameter p decreases. For this reason the parameter estimation for the GGoM curves varies the parameter p more than others.

Parameter estimation for GGoM						
CURVES	with GLM curves			with RM curves		
with p	r	b	p	r	b	p
1	2.3207	0.3237	1.0000	2.3516	0.3279	1.0000
0.995	2.3068	0.3220	1.0000	2.2635	0.3159	1.0000
0.99	2.2932	0.3201	1.0000	2.2761	0.3178	1.0000
0.98	2.1953	0.3068	1.0000	2.4303	0.3326	0.9948
0.95	1.9338	0.2712	1.0000	2.2822	0.3187	1.0000
0.85	1.2596	0.1678	0.9823	2.2986	0.3214	1.0000
0.8	1.2670	0.1500	0.9496	2.2716	0.3180	1.0000

Table 2.9: Experiment 4: parameter estimation for GGoM with GLM and RM data curves.

2.3.6 Experiment 4: empirical directed distances from the generalized Gompertz model (GGoM) to other models

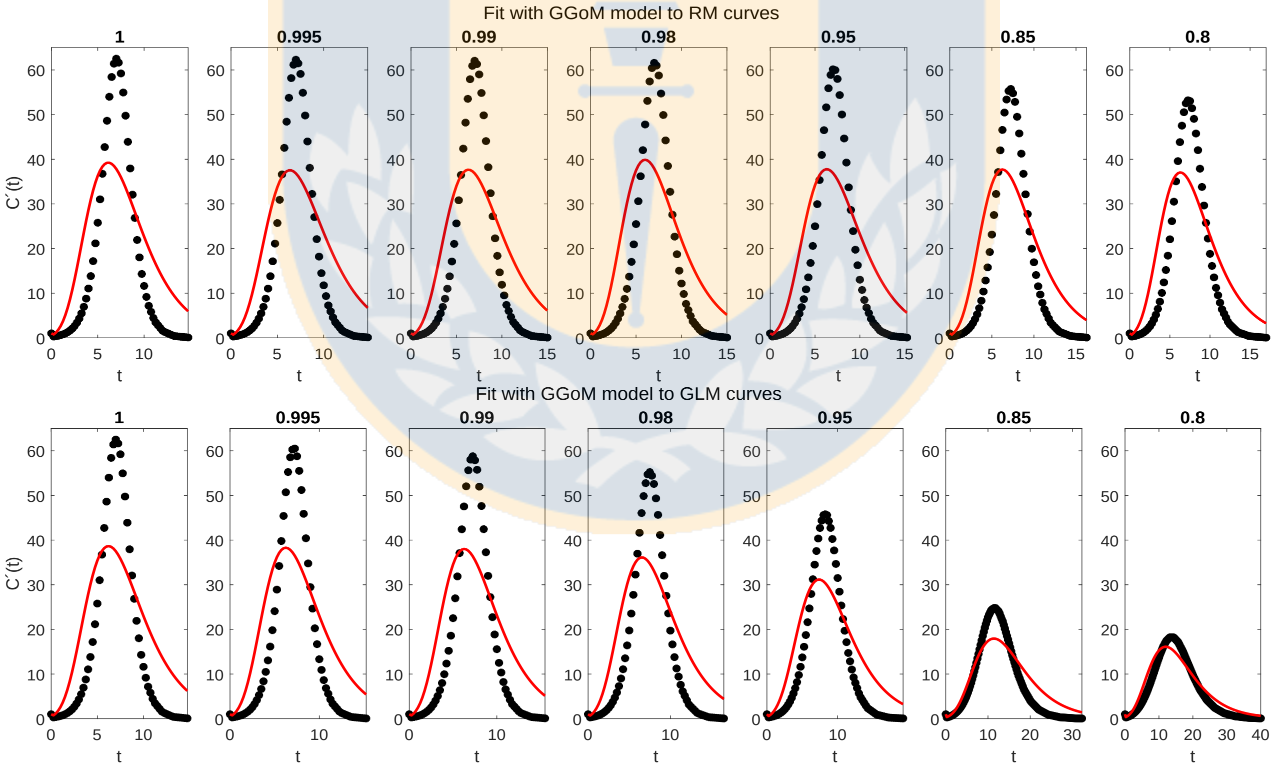


Figure 2.8: Experiment 4: results of fits of the GGoM (model B) to the curves of data generated by the RM (top row) and the GLM (bottom row), for the indicated values of p.

For this experiment, we consider the GGoM model as model B, and the models A are RM and GLM, with the parameters summarized in Table 2.2. In Figure 2.8 we can see the fits for RM and GLM data curves. This figure indicates that the GGoM does not capture the dynamics of the logistic models, where the maximum values are very large for the period of time defined in these data curves. The RMSEs in Table 2.5 are very large if compared with the previous experiments. The errors decrease when the parameter p is decreased, but this situation is due to approximation between the maximum values of the data curves and the maximum value that the GGoM can reach with the given period of time.

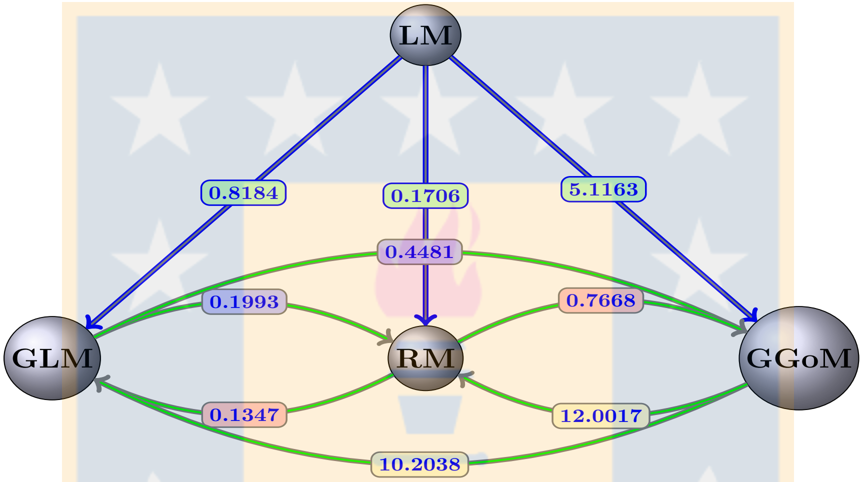


Figure 2.9: Comparative graph for each EDD and model.

Finally, the parameter estimation obtained for each fit is summarized in Table 2.9, where we observe that the parameter p is almost fixed. Being for the RM curves the other parameters almost equally fixed, this is due to the slow decrease for the maximum value. This contrasts with the result for the GLM, where the maximum value decreases faster than for the RM. For this reason the parameters r and b are varying. Summarizing, we have in Figure 2.9 the distances presented among the models studied, where each arrow indicates the direction of the distance from model B to model A.

	Influenza		Ebola		COVID-19	
Model	Interpolation	No interpolation	Interpolation	No interpolation	Interpolation	No interpolation
LM	28.4864	55.5662	44.5472	90.0674	59.6479	108.9944
GLM	26.2694	49.7601	24.4430	47.8958	21.2940	45.4623
GGoM	52.7972	113.9227	51.8345	94.6347	174.9620	356.6672
RM	29.1628	55.8399	26.2160	53.3493	54.9620	109.9705

Table 2.10: Application to real data: RSME for different time refinements.

	Influenza	Ebola	COVID-19
Model		Interpolation	
LM	$(0.5561, 2467.9)$	$(0.3141, 8988.2)$	$(0.3413, 9074.6586)$
GLM	$(0.5964, 1, 2228.8)$	$(0.7481, 0.8546, 10989)$	$(3.6232, 0.6869, 12963.9057)$
GGoM	$(1.244, 0.1809, 1)$	$(1.0000, 0.0897, 0.9487)$	$(5.7818, 0.0989, 0.6709)$
RM	$(0.5603, 1, 2655.7)$	$(0.4189, 0.4273, 11057)$	$(0.4188, 0.6302, 9196.4353)$
		No interpolation	
LM	$(0.5565, 2475.9)$	$(0.3127, 8327.7)$	$(0.3426, 9844.7509)$
GLM	$(0.6003, 1, 2363.1)$	$(0.7640, 0.8515, 11212)$	$(2.7782, 0.7213, 12316.4258)$
GGoM	$(1.2434, 0.1800, 1)$	$(0.8134, 0.0968, 1.0000)$	$(5.0086, 0.1020, 0.6931)$
RM	$(0.55451, 2392.5)$	$(0.4326, 0.4000, 11698)$	$(0.4133, 0.6472, 9145.3252)$

Table 2.11: Application to real data: parameter estimation for fit with real data

2.4 Examples: application to real data

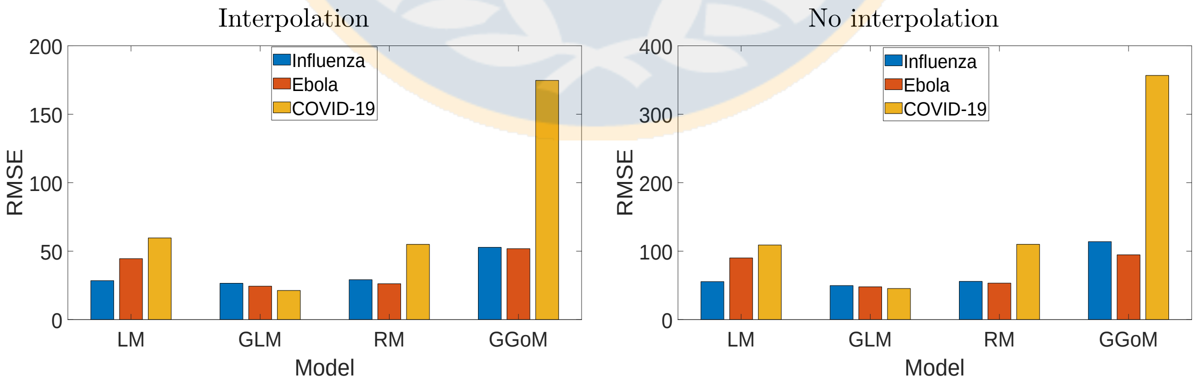


Figure 2.10: Application to real data: bar charts for the RMSE for each real data and refinement time.

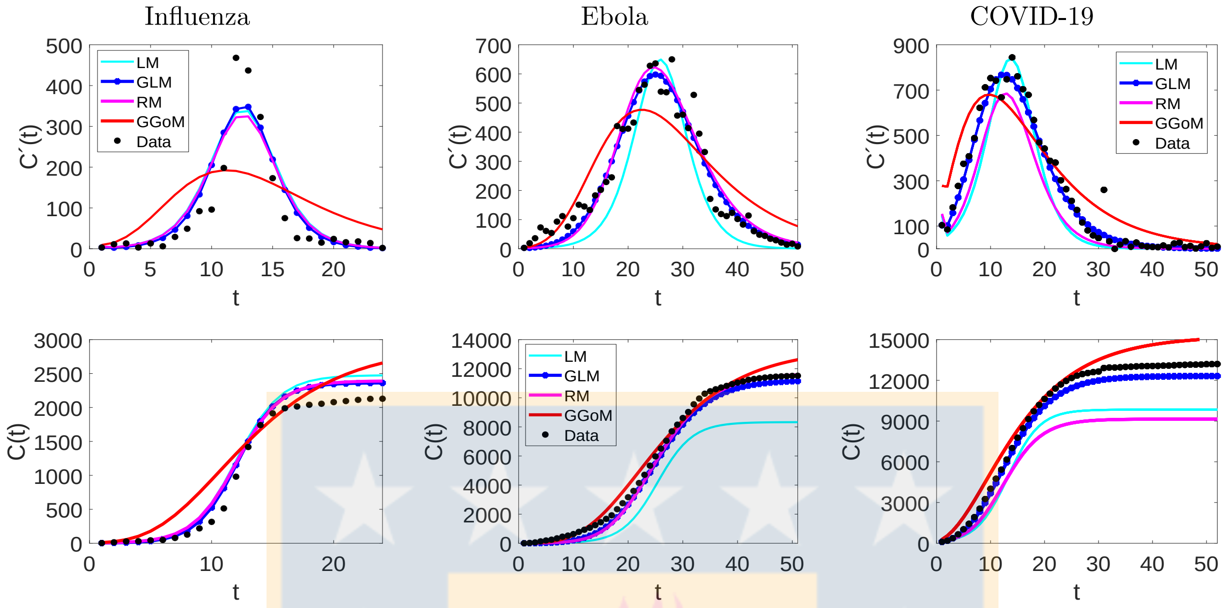


Figure 2.11: Application to raw data: fits to influenza, Ebola and COVID-19 data.

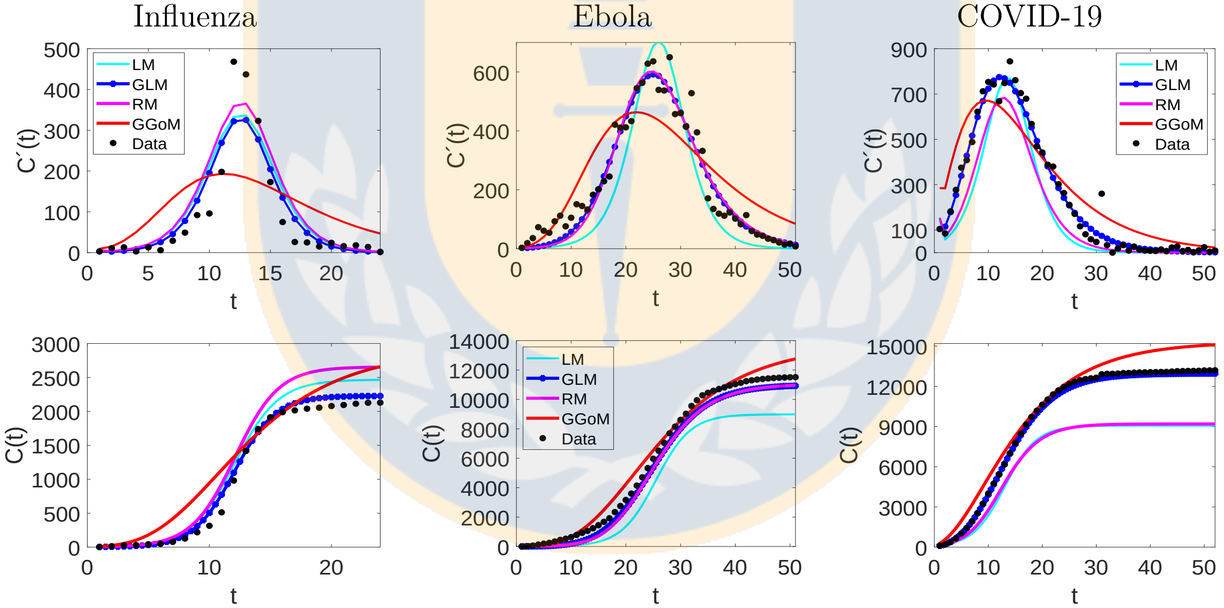


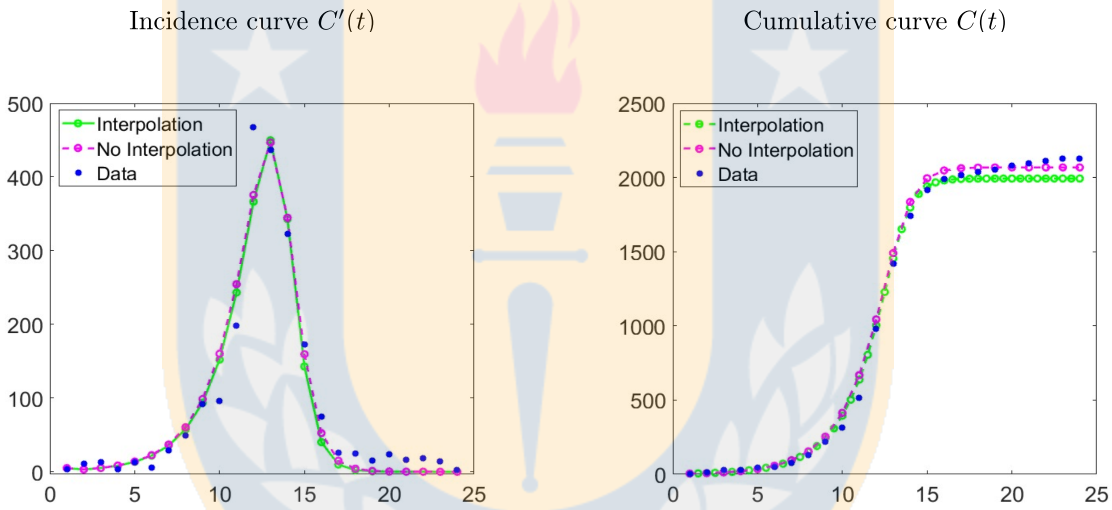
Figure 2.12: Application to interpolated data: fits to influenza, Ebola and COVID-19 data.

In order to see the best performance evidenced by the GLM model when capturing the other dynamics studied in the experiments performed, we present three examples with real data. In this case, we consider the data of weekly cases of influenza in Chile (24 data points in total) produced between autumn and winter of 2009 [59], Ebola (51 data points in total) in Sierra Leone dating from 2014 [118] and recent outbreak of COVID-19 [53] presented in various provinces of China (excluding Hubei province) (52 data points in total). Since we consider real data, for the application of the procedure of Section 2.3 we replace model A by real data but keep employing the same methodology of Section 2.3 with model B, where we also create a refinement of the real data by interpolation from the cumulative curve C, achieving for these examples twice the original number of points. From the RMSEs calculated and registered in Table 2.10 and the bar chart of Figure 2.10 we observe that the RMSEs for the non-interpolated data are close to the double from the RMSEs for the interpolated data, where effectively GLM meets be the best model with the smaller RMSE to the three examples.

From the figures of the fits, with and without interpolation (see Figures 2.11 and 2.12) for three examples, we can observe that the refinement from real data does not have a great impact on the performance of the GGoM (red), but in the Ebola case this model for early growth produces a better fit than others. For the fits made with the LM, we can observe that for the case of influenza the refinement leaves the fit similar to a fit without interpolation where for this case, the LM is better than the RM. A different situation occurs for the Ebola and COVID-19 cases where for Ebola the maximum value for the incidence curve increases and the cumulative curve increases close to real cumulative curve, though this is not better than the fits by the GLM and the RM. For COVID-19 the LM decreases the maximum value for the incidence curve and the cumulative curve decreases close to the cumulative curve of the RM, although this is not better than the fits by the GLM and RM. Now if we observe the fits with the RM and the GLM, we see that their fits though very similar for Ebola data, the GLM fits are better where the RMSE is smaller. On the other hand, with influenza data, we can see that for RM and GLM models, the curve with GLM is above the RM curve, staying in the middle the LM curve, and the situation changes when the data are interpolated, where the RM curve turns out to be above the GLM and LM curves, but the GLM produces the best fit with the smallest RMSE. In the case of COVID-19, the fits with the GLM with and without interpolated data are very close. A different situation occurs with the RM where the fits to the interpolated and non-interpolated data are below the data and therefore with RMSEs bigger than those for the GLM. Furthermore, Table 2.11 indicates that for the parameter estimation the values are very close between the real data and interpolated data, where for the LM this shows smaller variations and the GGoM model shows more variations with Ebola data.

Results	Influenza	Ebola	COVID-19
	Interpolation		
RMSE	17.2332	26.2223	$46,9405$
Parameter Estimation	$(0.4883, 3, 1993.2)$	$(0.4173, 0.4308, 11036)$	$(0.4551, 0.5408, 9895.1873)$
	No interpolation		
RMSE	28.7735	51.6678	$100,3531$
Parameter Estimation	$(0.49, 2.6641, 2068.1)$	$(0.4228, 0.4191, 11182)$	$(0.4350, 0.5877, 9556.1028)$

Table 2.12: Application to real data: results for different time refinements and real data for RM model with $p > 1$.



Incidence curve $C'(t)$

Cumulative curve $C(t)$

Figure 2.13: Application to real data: fits with Richards Model ($p > 1$) for influenza data.

CHAPTER 3

Modeling and forecasting of the transmission dynamics of the coronavirus 2019 pandemic in Colombia during 2020-2021

This chapter contributes to the study of transmission dynamics of SARS-Cov-2 in Colombia, using phenomenological growth models (PGMs) with case incidence and mortality data to estimate the potential transmission parameters and perform short-term forecasts of this epidemic's trajectory in Colombia. This chapter concentrates on the contributions presented in the work [150], which also gathers the contributions of other colleagues.

3.1 Introduction

We recall that the coronavirus disease 2019 (COVID-19) pandemic continues to threaten the world [125]. The non-pharmaceutical public health measures, including social distancing mandates and intermittent lockdowns, have been the principal strategies applied to fight COVID-19 [140]. Moreover, the rapid evolution of the SARS-CoV-2 virus contribute to the emergence of new variants amidst vaccination campaigns globally, making it more unclear how the COVID-19 pandemic will unfold [24, 138]. Therefore, the mortality and morbidity of the COVID-19 pandemic continue to overwhelm the health care systems of many nations, including the United States, Colombia, Mexico, Brazil, and Argentina in the Americas [64]

The pandemic emergency response, which includes the implementation and lifting of dynamic lockdowns [137], social distancing, mask mandates, population testing, and provision of vaccinations, has varied across countries in the same region and different cities within the same country [146]. For example, in Latin America, which is one of the epicenters of the pandemic [167], the evolution of the pandemic and the quality of public health responses have been different in each country. Colombia, despite aiming to contain virus transmission in the country, became the second country in Latin America and eighth globally to reach one million cases by the end of October 2020 [123]. Colombia has also seen the fastest increase in total

COVID-19 associated deaths compared to other Latin American countries, as reported by the end of the year 2020 [9].

Since the first confirmed case on March 6, 2020, Colombia has observed three epidemic peaks. On March 12, 2020 a national emergency status was declared, and educational institutions were closed by March 16, 2020, followed by the closure of the country's borders the next day [57,167].

Rappidly on March 23, 2020, as cases continued to increase rapidly, the domestic and international flights were suspended, and the next day, and a national quarantine was declared, to limit virus transmission within the country. Meanwhile, a phased reopening of the economy was initiated as early as April 27, 2020, under strict protocols to support the country's declining economy [4].

The evolution of the pandemic in Colombia has justified the five extensions of the mandatory quarantine that lasted until August 31, 2020 [160]. At this point, the country transitioned into a period of *selective isolation with responsible individual distancing* as the daily incidence in the country's main cities, including Bogotá, Medellín, Cali, Bucaramanga, and Pasto, leveled off and eventually leaned towards a downward trend [52, 99]. Moreover, Barranquilla, Cartagena, Leticia, and Quibdó had overcome the worst part of the first wave by August 25, 2020 [99]. The chosen selective isolation strategy prioritized tracing suspected cases, those with infection, and their contacts while reactivating the social and economic life of the country. As the cases continued to increase in the subsequent months, the government imposed and lifted dynamic lockdowns in multiple cities. By the end of the year 2020, COVID-19 cases were mainly concentrated in Bogotá, the capital city of Colombia, followed by the Antioquia and Valle del Cauca departments. [52].

The government announced a mass vaccination strategy that began on February 20, 2021 [155], and on February 25, 2021, an extension until May 31, 2021, of the national health emergency was declared, [122]. From March to June 2021, Colombia experienced a massive third wave of the COVID-19 pandemic, which evolved into a two-stage peak-within-a -peak surge [143]. The month of May 2021 was reported to be the deadliest month resulting in an average of 20000 cases and 500 deaths per day [83]. Bogotá, Antioquia, and Valle del Cauca have been the hardest-hit areas in Colombia [80] which is evidence that the impact of the COVID-19 pandemic was not uniform across the entire country.

Colombia is one of the first countries in Latin America to offer diagnostic tests for COVID-19 [51]; the testing and vaccination rate for Colombians remains low [95], with 0.79 tests per 1000 people per day and 0.39 vaccine doses administered per 100 people as of October 31, 2021. [186]. The factors contributing to the current COVID-19 outbreak are countless. However, three different events have particularly interacted synergistically to add to the complexity of the pandemic in Colombia. These include the COVID-19 outbreaks in prisons and nursing homes that affected the vulnerable communities of the society [72, 124] and the April 2021 Colombian protests provoked by the government policy proposals [143]. The COVID-19 pandemic severely impacted the Colombian prisons in Cali, Villavicencio, and Bogotá due to overcrowding, inadequate medical supplies, and unhygienic conditions of facilities, which led to many infected inmates [72]

The COVID-19 pandemic in Colombia presents complex risk dynamics of SARS-CoV-2 transmission with a simultaneous interplay of epidemiological, behavioral, and political factors. Forecasting the COVID-19 trajectory can help understand the disease trends and estimate its potential burden. As the epidemic trajectory of the COVID-19 pandemic continues to unfold, we forecast the COVID-19 course in near-real-time utilizing the mathematical models that have been validated for previous infectious disease outbreaks such as Ebola, Zika, and the COVID-19 pandemic [36, 119, 151, 153]. Moreover, we specifically investigate the transmission dynamics of SARS-CoV-2 at the national and regional levels, using the compute of the effective reproduction number, which allows us to observe the impact of the different control measures and social dynamics that occurred in Colombia during the 2020 and 2021.

3.2 Phenomenological Growth Models (PGMs)

In this contribution we highlight the generalized growth model (GGM), the generalized logistic growth model (GLM), Richards model (RM) and the novel sub-epidemic model. The first two models incoporate a parameter p that indicates the kind of scaling of growth. These models can be describe as follows;

The generalized growth model (GGM) relieson two parameters to characterize the early ascending phase of an epidemic where the model is given by the differential equation

$$\frac{\mathrm{d}C(t)}{\mathrm{d}t} = C'(t) = rC(t)^p,$$

where t is the time, C' describes the incidence curve over time, C is the cumulative number of cases at time while $r > 0$ indicates the intrinsic growth rate and $p \in [0, 1]$ is the modulating deceleration of growth parameter, where $p = 0$ correspond to constant incidence over time, $p = 1$ correspond to the exponential growth and the model shows sub-exponential growth dynamics if p is in the range $0 < p < 1$ [32, 170].

Similarly, the generalized logistic growth model (GLM) [142] incorporates the parameter $p \in [0, 1]$ to displays a range of epidemic growth patterns including the polynomial and exponental growth patterns. But GLM is defined by three parameters, the intrinsic growth rate $r > 0$, the growth scaling parameter, ($p \in [0, 1]$) and the final epidemic size K_0. During the initial stages of disease propagation, when $C(t) \ll K_0$ and $p = 1$ this model assumes the simple logistic growth model. The following differential equation gives the GLM model

$$\frac{\mathrm{d}C}{\mathrm{d}t}(t) = rC(t)^p \left(1 - \frac{C(t)}{K_0}\right). \tag{3.1}$$

where $C(t)$ and $C'(t)$ have the same meaning as for the GGM.

The well-known Richards model (RM) [127] is a extension of the logistic model that relies on three parameters, the growth rate, $r > 0$, the final epidemic size, K_0 and the scaling parameter, a, which measures the deviation from the symmetric S- shaped dynamics shown by the simple logistic growth curve [16,32,127,180]. The Richards model is given by the differential equation:

$$\frac{\mathrm{d}C\,(t)}{\mathrm{d}t} = rC(t)\left[1 - \left(\frac{C(t)}{K_0}\right)^{a}\right]$$

being $C'(t)$ like in the previous models. We remark that the Richards growth model has the explicit solution 2.11,

$$C(t) = \frac{K_0 C(0)\exp(rt)}{(K_0^a + C^a(0)(\exp(art) - 1))^{1/a}},$$

while the GLM does not admit a close-form solution. For a unified treatment of all phenomenological growth models (PGMs), we always refer to the corresponding differential equation in each case irrespective of the existence of a closed-form solution. Details are provided in a prior study [16].

Finally, the sub-epidemic model [44] is based on the premise that various profiles of overlapping sub-epidemics shape the aggregated reported epidemic wave. In particular, this modeling approach supports complex temporal dynamic patterns, such as oscillating dynamics leading to damped oscillations or endemic states. This model characterizes each group sub-epidemic by a three-parameter generalized logistic growth model as explained above and given in equation 3.1.

An epidemic wave comprising of n overlapping sub-epidemics is modeled using a system of coupled differential equations, as follows,

$$\frac{\mathrm{d}C_i(t)}{\mathrm{d}t} = \ rA_{i-1}\,(t)\,C_i(t)^p\left(1 - \frac{C_i(t)}{K_i}\right)$$

where $C_i(t)$ describes the cumulative cases for i-th sub-epidemic, and K_i is the size of sub-epidemic $i = 1 \ldots .n$. Parameters r and p are the same across the sub-epidemics. Therefore, when $n = 1$ and $p = 1$, the sub-epidemic model becomes the simple logistic model. $A_i(t)$ is an indicator variable that models the onset timing of $(i + 1)$-th sub-epidemic, making sure that sub-epidemics comprising an epidemic wave follow a regular structure.

Therefore,

$$A_i\,(t) = \begin{cases} 1, & C_i(t) > C_{thr} \\ 0, & \text{Otherwise} \end{cases} \quad i = 1, 2, 3, \ldots n \tag{3.2}$$

with $1 \leq C_{thr} < K_0$ and $A_1\,(t) = 1$ for the first sub-epidemic.

Moreover, for the subsequently occurring sub-epidemics, the size of i-th sub-epidemic (K_i) declines exponentially at a rate q, i.e.,

$$K_i = K_0 \exp(-q(i - 1)),$$

Growth model	**Parameters**
Generalized growth model (GGM)	$\Theta = \{\vartheta_1 = r, \vartheta_2 = p\}$
Generalized Logistic growth model (GLM)	$\Theta = \{\vartheta_1 = r, \vartheta_2 = p, \vartheta_3 = K_0\}$
Richards model (RM)	$\Theta = \{\vartheta_1 = r, \vartheta_2 = a, \vartheta_3 = K_0\}$
Sub-epidemic model	$\Theta = \{\vartheta_1 = r, \vartheta_2 = p, \vartheta_3 = K_0, \vartheta_4 = q, \vartheta_5 = C_{thr}\}$

Table 3.1: Summary of information about models and parameters. Being, $r, p, q, K_0 > 0,\ p \in [0, 1]$ and $1 \leq C_{thr} < K_0$.

with K_0 the initial sub-epidemic. The exponential decline in the size of the i-th sub-epidemic can occur due to multiple factors, including the effect of interventions, changes in disease transmission dependent on seasonality and behavior changes [44].

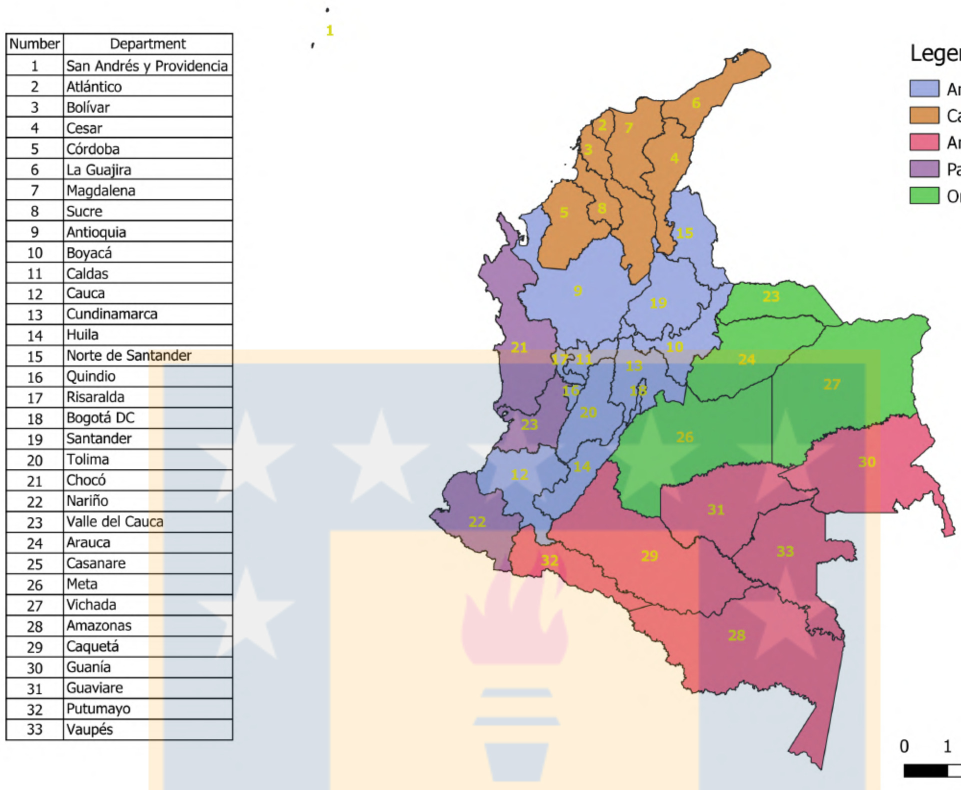
These models have been applied to various infectious diseases including SARS, foot and mouth disease, Ebola [22, 44, 142] and the current COVID-19 outbreak [16, 109, 149]. For our study we utilize these PGMs to fit data and to generate short-term (i.e., 30-day ahead) forecasts for Colombia. The forecasts obtained from these dynamic growth models can assess the potential scope of the pandemic in near real-time, provide insights on the contribution of disease transmission pathways, predict the impact of control interventions and evaluate optimal resource allocation to inform public health policies. The following section summarize the material and methods used to study the COVID-19 epidemic in Colombia, such as the data, the performance metrics, the calibration process, among other topics.

3.3 Materials and methods

To understand the transmission dynamics of COVID-19 in Colombia, we need time series data, we herein employ the daily incidence case and death data. Adittionally, we will apply the models described in Section 3.2 to fit and calibrate regional and national data in different periods of the epidemic. Then with these results, we evaluate the best fits and diverse short-term forecasts generated with the same models using the performance metrics that will be defined in this section, together with the methodology of fitting, calibration, and forecasting.

3.3.1 Data

We wish to obtain information about the effectiveness of intervention strategies applied in Colombia and the effect of other social dynamics. Then, we decided to use two types of data, the case incidence data and mortality data, both retrieved from the Colombian Ministery of Health as of October 31, 2021 [52]. Specifically, the case incidence data based on symptom onset date is used to generate the epidemic curve, the short-term forecasts, and estimate the national and



Number	Department
1	San Andrés y Providencia
2	Atlántico
3	Bolívar
4	Cesar
5	Córdoba
6	La Guajira
7	Magdalena
8	Sucre
9	Antioquia
10	Boyacá
11	Caldas
12	Cauca
13	Cundinamarca
14	Huila
15	Norte de Santander
16	Quindío
17	Risaralda
18	Bogotá DC
19	Santander
20	Tolima
21	Chocó
22	Nariño
23	Valle del Cauca
24	Arauca
25	Casanare
26	Meta
27	Vichada
28	Amazonas
29	Caquetá
30	Guanía
31	Guaviare
32	Putumayo
33	Vaupés

Figure 3.1: Geographical Colombia distribution consists of 32 departments and a capital district, listed and referenced with a number on the Colombian map. Additionally, the departments from each Colombia region are identified with colors. Map edited by the author and used as Supporting information of work [150] `https://doi.org/10.1371/journal.pntd.0010228.s005`.

regional reproduction numbers. Additionally, the mortality data based on the date of death is used to generate a national short-term forecast and estimate the national reproduction number. Details of the geographical distribution of the Colombian regions see the map in Figure 3.1.

3.3.2 Performance metrics

With the performance metrics, we can quantify the error of the model fit to the data, such as in mentioned in [89] and is applied in [32]. Then we incorporate five performance metrics to assess the quality of our model fit and the 30-day ahead short-term forecasts. These are the mean absolute error (MAE), the root mean squared error (RMSE), the coverage of the 95% prediction intervals (95% PI), the mean interval score (MIS), and the weighted score (WIS) (more details see [76]). Below is a summary of each metric.

The root mean squared error (RMSE) and the mean absolute error (MAE) assess the average deviations of the model fit to the observed data. The root mean squared error (RMSE) is given by

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f(t_i,\hat{\Theta}) - y_{t_i})^2},$$

and the mean absolute error (MAE) is given by

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|f(t_i,\hat{\Theta}) - y_{t_i}|. \tag{3.3}$$

In both these cases equations $\hat{\Theta}$ s the set of parameter estimated, $f(t_i,\hat{\Theta})$ denotes the best-fit model, and y_{t_i} $(i = 1,\ldots,n)$ is the time series of cases by date of onset, t_i is the time stamp and n is the total number of data points, for the calibration period, and for the forecasting period, $n = 30$ for the 30-day ahead short-term forecast.

Moreover, to assess the model uncertainty and performance of prediction interval, we use the coverage of the 95% prediction intervals (95% PI), the mean interval score (MIS), and the weighted score (WIS). The prediction coverage is defined as the proportion of observations that fall within 95% PI, and it is calculated as,

$$\text{PI coverage} = \frac{1}{n}\sum_{i=1}^{n} I\left\{y_{t_i} > L_{t_i} \ \cap \ y_{t_i} < U_{t_i}\right\}$$

where y_{t_i} are the case incidence data,L_{t_i} and U_{t_i} are the lower and upper bounds of the 95% PIs, respectively, n is the length of the period, and I is an indicator variable that equals 1 if value of y_{t_i} is in the specified interval and 0 otherwise

The MIS addresses the width of the PI as well as the coverage, and it is given by,

$$\mathrm{MIS} = \frac{1}{n}\sum_{i=1}^{n}(U_{t_i} - L_{t_i}) + \frac{2}{0.05}(L_{t_i} - y_{t_i})I\left\{y_{t_i} < L_{t_i}\right\} + \frac{2}{0.05}\ (U_{t_i} - y_{t_i})\, I\left\{y_{t_i} > U_{t_i}\right\},$$

in this equation L_{t_i}, U_{t_i}, y_{t_i}, n and I are as specified above for PI coverage. Therefore, if the PI coverage is 1, the MIS is the average width of the interval across each time point. For two models that have an equivalent PI coverage, a lower value of MIS indicates narrower intervals [76].

Finally, the Weighted interval score (WIS) is a proper score that provides quantiles of predictive forecast distribution by combining a set of interval scores (IS) for probabilistic forecasts. An interval score is simple proper score requires only a central $(1-\alpha)\times 100\%$ prediction interval (PI) [76] and is described as

$$\mathrm{IS}_\alpha(F,y) = (u-l) + \frac{2}{\alpha}(l-y)I\,(y<1) + \frac{2}{\alpha}\,(y-u)\,I\,(y>u)$$

where I refers to the indicator function, meaning that $I\,(y<1) = 1$ if $y < l$ and 0 otherwise. The terms l and u represent the $(\alpha/2)$ and $(1-\alpha/2)$ quantiles of F. The IS consists of three distinct quantities,

1. The sharpness of F, given by the width $u - l$ of the central $(1-\alpha)PI$.

2. A penalty term $\frac{2}{\alpha}(l-y)\ I(y<l)$ for the observations that fall below the lower end point l of the $(1-\alpha) \times 100\% PI$. This penalty terms directly proportional to the distance betweem the lower end l of the PI and y. The stregth of the penalty depends on the level α.

3. An analogous penalty term $\frac{2}{\alpha}(l-u)\ I(y>l)$ for all the observations that fall above the upper end u of the PI.

To provide more detailed and accurate information on the entire predictive distribution, we report several central PIs at different intervals $(1-\alpha_1) < (1-\alpha_2) < \cdots < (1-\alpha_k)$ along with the predictive median, m, which can be seen as a central prediction interval at level $1-\alpha_0 \to 0$. The WIS as a particular linear combination of K intervals scores is defined as follows,

$$\mathrm{WIS}_{\alpha_{0:K}}(F,y) = \frac{1}{K+\frac{1}{2}}\left(w_0|y-m| + \sum_{k=1}^{K}(w_k IS_{\alpha_k}(F,y))\right),$$

where $w_k = \frac{\alpha_k}{2}$ for $k = 1, 2, \ldots, K$, and $w_0 = \frac{1}{2}$. Hence, WIS can be interpreted as a measure of how close the entire distribution is to the observation, in units on the scale of the observed data [12, 56].

3.3.3 Model calibration and forecasting approach

We utilize the national and regional level case incidence data and national level mortality data to obtain the best-fits using each model detailed in Section 3.2. Specifically, each forecast is fitted to the daily case counts based on the dates of symptom onset and daily death counts based on the date of death between July 4, 2021 and October 1, 2021 (90 days calibration period) to conduct a 30-day ahead short-term forecast for each model. The data from October 2, 2020, to October 31, 2021, is utilized to assess the performance of our 30-day ahead short-term forecasts.

The best-fit solution for each model (i.e., the GLM, RM and sub-epidemic models) $f(t, \hat{\Theta})$ is obtained using a non-linear least squares fitting procedure [7]. This process yields the best set of parameter estimates $\hat{\Theta} = (\hat{\vartheta}_1, \hat{\vartheta}_2, \ldots, \hat{\vartheta}_m)$ (where m is the number of parameters of interest), that minimizes the sum of squared errors between the model fit, $f(t, \Theta)$ and the observed data, y_{t_i} $(i = 1, 2, \ldots, n)$. That is summarize in the objective funtion given by,

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \sum_{i=1}^{n}(f(t_i, \Theta) - y_{t_i})^2,$$

where t_i is the timestamps at which time series data are observed, and n is the number of data points available for inference.

Concretely in our study, we have the following parameter sets, $\Theta = (r, p, K_0)$ for the GLM model, $\Theta = (r, K_0, a)$ for the RM model and $\Theta = (r, p, K_0, q, C_{thr})$ for the sub-epidemic model, more details in Table 3.1.

For the sub-epidemic wave model, we determine the initial best guesses of parameter estimates. However, for the GLM and RM we initialize the parameter estimates for the nonlinear least squares method [7] over a wide range of plausible parameters from a uniform distribution using Latin hypercube sampling provided by Matlab fuction `lhsdesign()`. This allows us to test the uniqueness of the best model fit. The initial conditions are set at the first data point for each of the three models [32].

Using a parametric bootstrap approach with data replacement are generated uncertainty bounds around the best-fit solution. We assume a negative binomial error structure for the PGMs considered for the fitting process. For case incidence data, the variance (for the negative binomial error) is assumed to be 488.85 times the mean for national data, 11.59 times the mean for the Amazon region, 356.8 times the mean for the Andean region, 69.72 times the mean for the Caribbean region, 77.93 times the mean for the Pacific region, and 22.17 times of the mean for the Orinoquia region. The variance is assumed to be 17.95 times the mean for the mortality data. The variance is based on data noise and calculated by averaging the mean to variance ratio obtained from the data. A detailed description of this method is provided in a prior study [32].

From the parametric bootstrap approach are obtained $S = 300$ best-fit parameter sets, which are used to construct each parameter's 95% confidence intervals. Further, for each S best fit model solution, $s = 30$ additional simulations are generated with a negative binomial error structure for each PGM extended through a 30-day forecasting period. Finally, we construct the 95% prediction intervals with these $9000(S \times s)$ curves for the forecasting period. More details about the methods of parameter estimation and bootstrap approach can be found in references [32, 134, 136].

3.4 Reproduction number

The effective reproduction number R_t is the key parameter that characterizes the average number of secondary cases generated by a primary case at calendar time t during an outbreak. This quantity is crucial for identifying the magnitude of public health interventions required to contain an epidemic [2, 33, 114]. The estimates of R_t indicate whether widespread disease transmission continues ($R_t > 1$) or disease transmission declines ($R_t < 1$). Therefore, to contain an outbreak, it is vital to maintain $R_t < 1$.

In light of the properties of the effective reproduction number and our interest in understanding the transmission dynamics of COVID-19 in Colombia, we estimate it for two phases, the first in the early ascending phase between February 27, 2020, and March 27, 2020, and the other throughout the pandemic for the national and regional COVID-19 epidemic curves. A way was applied to compute this reproduction number, the GGM (as in [170]) is explained in the following subsection.

3.4.1 Effective Reproduction Number R_t using the GGM

The national and regional reproduction numbers are estimated by calibrating the GGM to the early growth phase of the pandemic [170]. We first characterize the daily incidence of local cases using the GGM. The progression of local incidence cases by dates of symptom onset, I_i, is simulated using the calibrated GGM model and accounts for the daily series of imported cases by dates of symptom onset, J_i, into the renewal equation to estimate the effective reproduction, R_{t_i} as

$$R_{t_i} = \frac{I_i}{\sum_{j=0}^{i} (I_{i-j} + \ \alpha J_{i-j})\rho_j},$$

where the factor J_i represents the imported cases at time t_i, I_i denotes the local case incidence at calendar time t_i, and ρ_j represents the discretized probability distribution of the generation interval, besides the factor $0 \leq \alpha \leq 1$ represents the relative contribution of imported cases to secondary disease transmission. Therefore, in the numerator, we have the total new cases I_i, and in the denominator, the total number of primary cases contributes to the generation of secondary cases I_i at time t_i. Hence, R_t represents the average number of secondary cases generated by a single case at calendar time t. Additionally, the generation interval of the SARS-Cov-2 virus is modeled with the assumption of a Gamma distribution with a mean of 5.2 days and a standard deviation of 1.72 days [148].

The uncertainty bounds around the curve of R_t are derived directly from the uncertainty associated with the parameter estimates from the GGM, and this uncertainty is generated using $S = 300$ bootstrapping samples from GLM with a negative binomial error structure and a variance three times the mean based on the noise of the data [32]. For each sample a parameter pair $(\hat{r}_i, \hat{p}_i)$ $i = 1, \ldots S$ is estimated. This method is utilized to derive the early estimates of reproduction numbers and has been employed in several prior studies [32,109,152].

Since the national and regional epidemic curves have a different onset date according to the first local case's generating, R_t for the national and regional level is estimated starting on the onset of the first local case. For the national epidemic curve, we estimate R_t for the first 30 days, following the distribution shown in Table 3.2.

Finally, we include a sensitivity analysis to assess the relative contribution of the imported cases to the secondary transmission [151], setting three values for $\alpha = \{0.00, 0.15, 1.00\}$ in the computation of R_t.

Region	Dates(2020) Month/DD
National	February 28 - March 28
Orinoquía region	March 17 - April 17
Amazon region	March 25 - April 23
Caribbean region	February 29 - March 29
Andean region	March 01 - March 30
Pacific region	February 28 - March 28

Table 3.2: Dates for R_t estimation for the national and regional epidemic curves for the first 30 epidemic days

3.5 Results

As of October 31, 2021, Colombia has reported 5.002.387 cases based on symptom onset dates. Andean region concentrates about 64% of total cases, followed by the Caribbean region with around 20%, the Pacific with 12%, the Orinoquia with 3% and finally the Amazon only with 1%. The COVID-19 epidemic curve in Colombia shows a five-modal pattern, with the first peak occurring in mid-July 2020 after the phased reopening of the country. In mid-October 2020, a slight surge occurred after selective isolation and social distancing interventions. The second peak occurred at the beginning of January 2021, and the third occurred from March through June 2021. While the mortality curve shows a three-modal pattern, the first peak occurred in July 2020, followed by a second peak in January 2021, and a third more prominent peak from April through June 2021, which coincide with the three epidemic waves observed in the case incidence data (Information illustrated in Figures 2 and S2 of [150]).

A timeline showing the major events during the COVID-19 pandemic in Colombia is presented in 3.2.

3.5.1 Model calibration and forecasting performance

We compare the results from model calibration by 90 epidemic days and 30-day ahead forecasting across the GLM, RM and the sub-epidemic wave model. Model calibration, using 90 days of the epidemic data, between July 4, 2021 and October 1, 2021 shows that at the regional and national levels, the sub-epidemic model outperformed the GLM and RM in terms of the all five performance metrics (see Table 3.3) using the case incidence data. Therefore, the sub-epidemic wave model can be declared the most accurate model for the calibration period. The model fits exhibited sub-exponential growth dynamics [22,170] for three models ($p \sim 0.6 - 0.8$) at the national and regional levels. The calibration performances for each region and the national data are listed in Table 3.3. Calibrating the three models to the mortality data also

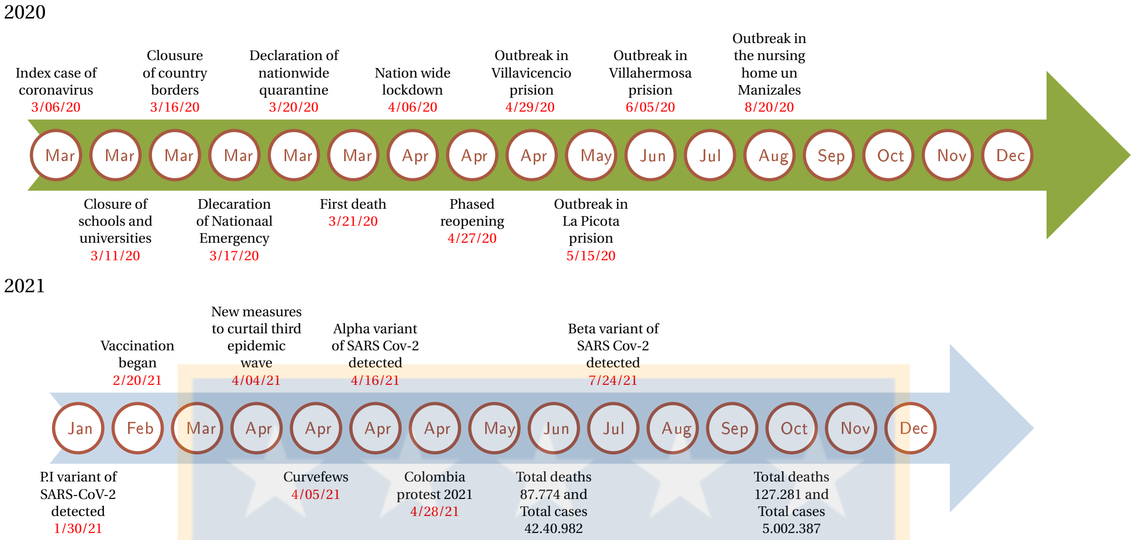

2020
Index case of coronavirus 3/06/20
Clousure of country borders 3/16/20
Declaration of nationwide quarantine 3/20/20
Nation wide lockdown 4/06/20
Outbreak in Villavicencio prision 4/29/20
Outbreak in Villahermosa prision 6/05/20
Outbreak in the nursing home un Manizales 8/20/20
Mar
Mar
Mar
Mar
Mar
Mar
Apr
Apr
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Closure of schools and universities 3/11/20
Dlecaration of Nationaal Emergency 3/17/20
First death 3/21/20
Phased reopening 4/27/20
Outbreak in La Picota prision 5/15/20
2021
Vaccination began 2/20/21
New measures to curtail third epidemic wave 4/04/21
Alpha variant of SARS Cov-2 detected 4/16/21
Beta variant of SARS Cov-2 detected 7/24/21
Jan
Feb
Mar
Apr
Apr
Apr
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
P.I variant of SARS-CoV-2 detected 1/30/21
Curvefews 4/05/21
Colombia protest 2021 4/28/21
Total deaths 87.774 and Total cases 42.40.982
Total deaths 127.281 and Total cases 5.002.387

Figure 3.2: Timeline of the COVID-19 pandemic in Colombia as of October 31, 2021. Figure adapted from `https://doi.org/10.1371/journal.pntd.0010228.g001` in [150]

shows that the sub-epidemic model outperforms the other two models (Table B.1) and exhibits sub-exponential growth dynamics.

In terms of the forecasting performance, again, the sub-epidemic wave model performed better than the GLM and RM for the national and regional case incidence data (Table 3.4) and the mortality curve (Table B.2). Hence, the sub-epidemic model can be considered the most accurate model to forecast the epidemic trajectory for incidence and mortality cases.

3.5.2 30-day ahead forecasts

Calibrating our models from July 04 to October 1, 2021, and generating the 30-day ahead forecasts from October 2 to October 31, 2021, utilizing the GLM and RM indicates a downwards trend for the national and regional case incidence data, as is shown in Figures 3.3 and 3.4). On the other hand, the sub-epidemic wave model captures the multiple sub-epidemics comprising the course COVID-19 epidemic wave of Colombia. The sub-epidemic model predicts a downward trend for the Amazon and Orinoquía region, consistent with the findings of the GLM and RM. While, for the national, Andean, Caribbean, and the Pacific region, the sub-epidemic model predicts a stable case incidence pattern as evidenced in Figure 3.5.

According to the forecasts obtained by GLM and RM, the COVID-19 pandemic in Colombia would decline to zero during the month of October while the sub-epidemic wave model predicts an accumulation of 24525 (95% PI:$13677, 44515$) cases at the national level for the month of October, 2021.

At the regional level, RM and GLM predict zero cases for the month of October (Figures 3.3-

Region	RMSE	MAE	MIS	95% PI	WIS
	GLM				
National	884.72	1315.3	18776	47.78	993.90
Pacific	97.36	152.17	1970.40	56.67	116.65
Caribbean	166.11	265.68	4457.80	61.10	207.89
Andean	467.69	738.12	7797.10	61.11	542.25
Amazon	1.94	11.70	86.04	98.89	8.15
Orinoquía	19.13	34.25	338.41	57.78	25.49
	Richards model				
National	1236.7	1643.5	31087	31.11	1323.2
Pacific	166.66	221.63	4186.9	31.11	180.51
Caribbean	300.60	372.66	9517.4	25.56	320.68
Andean	714.72	989.42	16221	37.78	771.77
Amazon	12.75	21.50	294.10	44.44	16.02
Orinoquía	40.90	53.87	1231.6	16.67	45.58
	Sub-epidemic wave model				
National	442.6	298.15	1941.1	98.88	187.03
Pacific	93.18	62.48	314.25	96.67	93.18
Caribbean	134.47	99.02	666.6	91.11	64.98
Andean	353.9	219.76	12710	98.88	143.88
Amazon	18.46	10.61	85.39	95.55	6.87
Orinoquía	36.42	19.300	145.06	94.44	12.86

Table 3.3: Performance metrics by calibrating the GLM, RM and the sub-epidemic model for 90 epidemic days (July 4, 2021 to October 1, 2021) at the national and regional level. Higher 95% PI coverage and lower RMSE, MAE, WIS and MIS represent better performance. We highlight best performing model with green color.

3.4). However, the sub-epidemic model predicts 59 (95% PI: $0, 507$) cases for the Amazon region in the month of October. The sub-epidemic model also predicts 12186 (95% PI:$7107, 19235$) cases for the Caribbean region, 18165 (95% PI:$9205, 31205$) cases for the Andean region, 5039 (95% PI:$2517, 8725$) cases for the Pacific region and 510 (95% PI:$87, 1456$) cases for the Orinoquía region (Figure 3.5). The 30-day ahead forecast of the national mortality data generated by GLM and RM indicate a decline in the number of deaths, and the sub-epidemic wave model indicates an upward trend in the mortality curve with 2893 (95% PI:$1860, 5325$) deaths that can accumulate in the month of October 2021 (Figure B.1).

Region	RMSE	MAE	MIS	95% PI	WIS
			GLM		
National	1328.1	1312.5	101300	0	1037.8
Pacific	144.09	141.42	10682	0	115
Caribbean	499.71	487.00	30368	61.1	279.17
Andean	655.34	646.37	43420	0	609
Amazon	13.62	12.39	337.39	17.24	4.24
Orinoquía	20.35	18.97	1537.6	0	17
			Richards model		
National	1322.8	1307.2	130740	0	1038
Pacific	143.84	141.25	16478	0	115
Caribbean	500.28	487.14	45484	0	219
Andean	655.14	646.79	63713	0	609
Amazon	14.16	13.03	1075.2	0	6
Orinoquía	20.29	18.99	3214.3	0	17
			Sub-epidemic wave model		
National	466.32	405.5	26356000	73.33	273.19
Pacific	41.746	36.03	210.01	100	20.44
Caribbean	135.15	105.5	676.2	86.67	69.68
Andean	109.38	90.69	733.3	100	142.9
Amazon	11.68	10.61	60.3	66.67	8
Orinoquía	7.266	5.80	45.6	100	3.90

Table 3.4: 30-day ahead forecasting performance (October 2, 2021 to October 31, 2021) by calibrating the GLM, RM and the sub-epidemic model for 90 epidemic days (July 4, 2021 to October 1, 2021) at the national and regional level. Higher 95% PI coverage and lower RMSE, MAE, WIS and MIS represent better performance. We highlight best performing model with green color.

3.6 Reproduction number

3.6.1 Estimate of the effective reproduction number, R_t from case incidence data

The reproduction number for the early ascending growth phase of the epidemic from the case incidence data (February 27, 2020 to March 27, 2020) using GGM was estimated at $R_t \sim 1.30$ (95% CI:$1.20, 1.50$) at $\alpha = 0.15$ for the national data. The growth rate parameter, r, was estimated at 1.40 (95% CI:$0.91, 2.0$) and the deceleration of growth parameter, p, was estimated at 0.64 (95% CI: $0.56, 0.71$), indicating early sub-exponential growth dynamics of the epidemic (see Figure 3.6). Simultaneously, the estimates of R_t for all the regions remained consistently above 1.2 (between $\sim 1.20 - 2.22$) for the early ascending phase of the pandemic with the

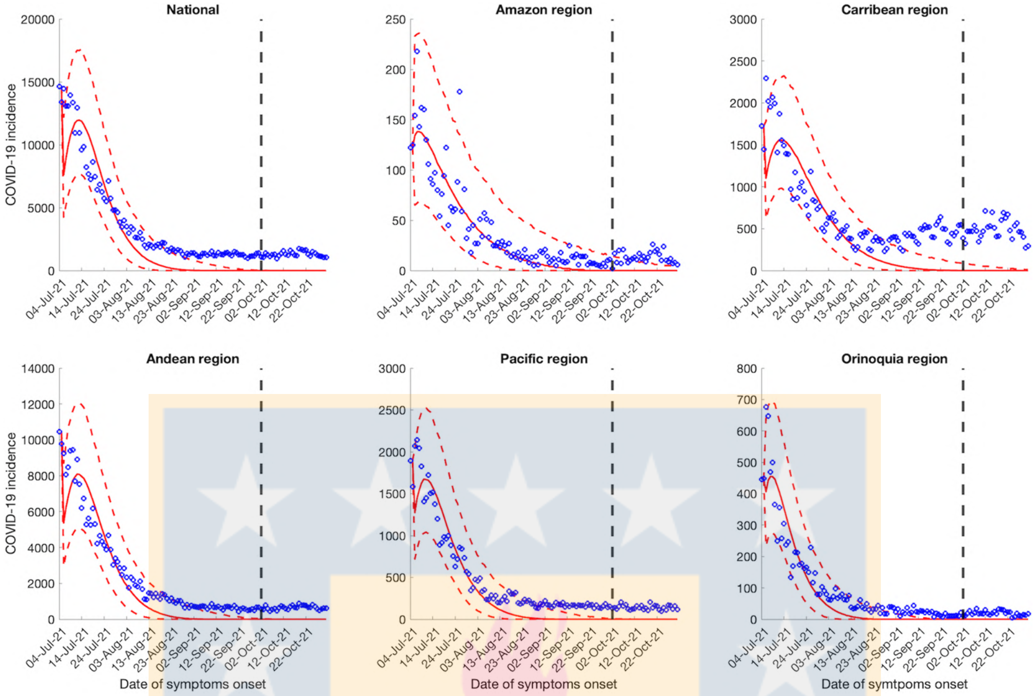


Figure 3.3: 30-days ahead forecasts of the national and regional COVID-19 epidemic curves in Colombia by calibrating the GLM model from July 4, 2021 to October 1, 2021. Here and Figures 3.4 and 3.4 the blue circles correspond to the data points; the solid red line indicates the best model fit, and the red dashed lines represent the 95% PI. The vertical black dashed line represents the time of the start of the forecast period. The figure is taken from the published article [150].

Amazon region showing the highest estimate of reproduction number, $R_t \sim 2.2$ followed by Orinoquía region with $R_t \sim 1.8$. The estimates of R_t for the Andean, Pacific and Caribbean regions remained between $R_t \sim 1.2 - 1.4$. All regions except the Amazon region depict sub-exponential growth dynamics for the COVID-19 pandemic in Colombia with the deceleration of growth parameter, p, estimated between $p \sim 0.54 - 0.86$. The Amazon region shows almost exponential growth dynamics with the deceleration of growth parameter, $p \sim 0.95$ (95% CI: 0.74, 1.00). In contrast, Andean and Caribbean regions show an almost linear pattern of the epidemic trajectory with the deceleration of growth parameter, $p \sim 0.58$ (95% CI:0.48- 0.69) and $p \sim 0.59$ (95% CI: 0.34, 0.87) respectively. The sensitivity analysis shows that the estimates of reproduction numbers do not vary significantly at $\alpha = 0.15$ and $\alpha = 1.00$ (Table 3.5). Moreover, setting $\alpha = 0.00$, thereby assuming that there is no contribution of imported cases to the disease transmission process or generation of secondary cases also does not substantially change the estimates of reproduction number.

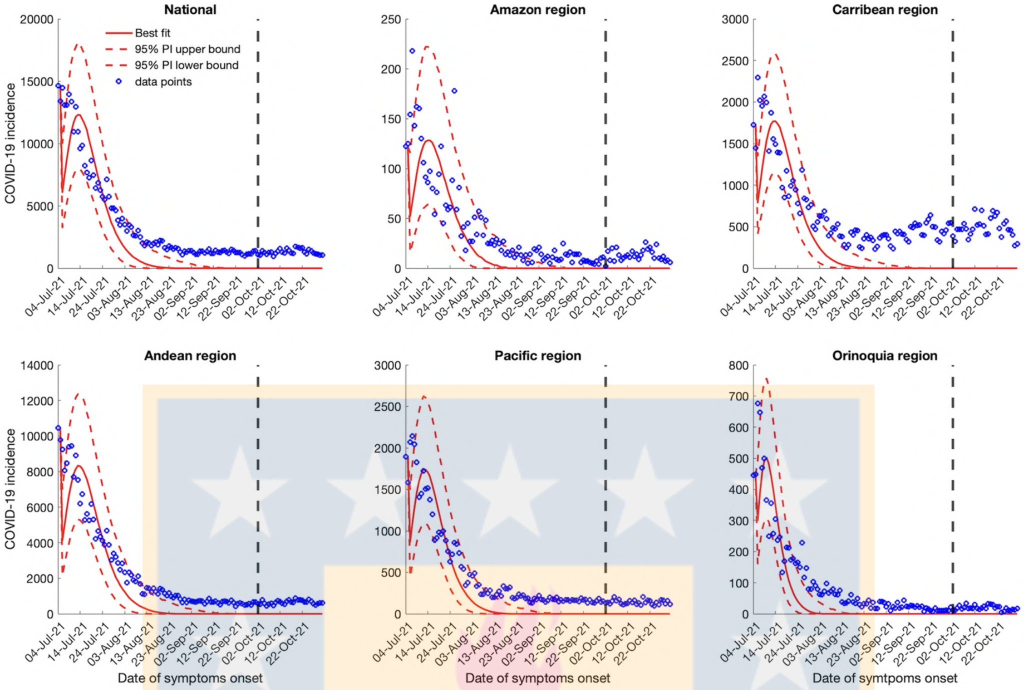


Figure 3.4: 30-days ahead forecast of the national and regional COVID-19 epidemic curves in Colombia by calibrating the Richards model from July 4, 2021 to October 1, 2021. The figure is taken from the published document. [150].

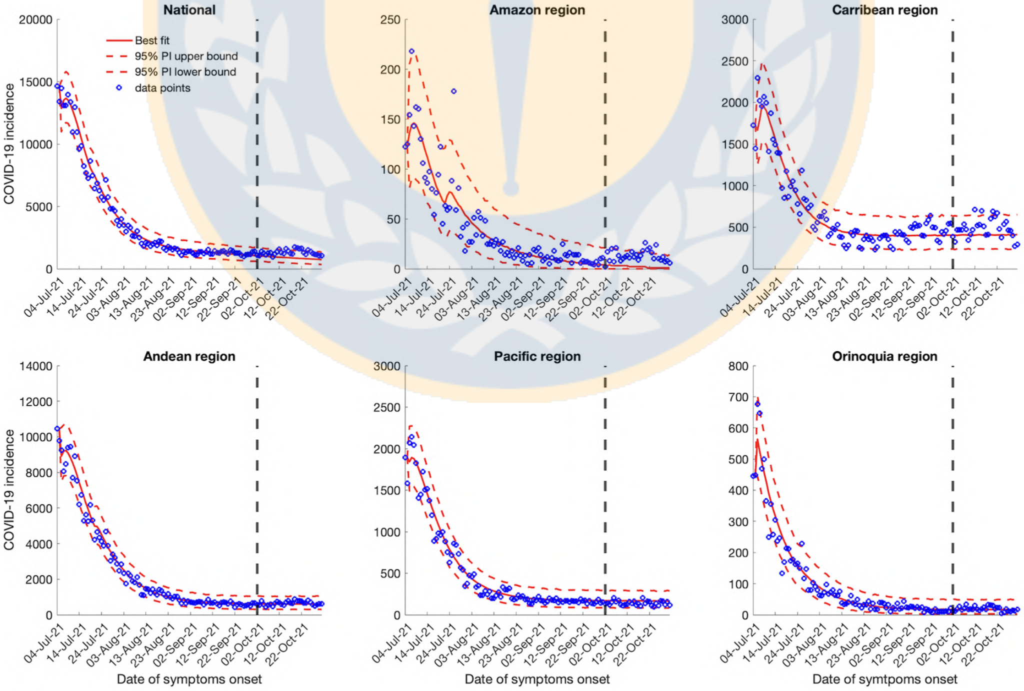


Figure 3.5: 30-days ahead forecast of the national and regional COVID-19 epidemic curves in Colombia by calibrating the sub-epidemic wave model from July 4, 2021 to October 1, 2021. The figure is taken from the published article [150].

Region	r		p		R	
	mean	$95\%CI$	mean	$95\%CI$	mean	$95\%CI$
			$\alpha = 0.15$			
National	1.40	(0.91,2.0)	0.64	(0.56,0.71)	1.3	(1.2,1.5)
Amazon	0.23	(0.18,0.4)	0.95	(0.74,1.0)	2.2	(1.5,2.6)
Andean	1.70	(0.88,2.8)	0.58	(0.48,0.69)	1.2	(1.1,1.4)
Caribbean	0.85	(0.27,1.90)	0.59	(0.34,0.87)	1.3	(1.1,1.8)
Pacific	0.63	(0.27,1.9)	0.69	(0.48,0.9)	1.4	(1.1,2.0))
Orinoquia	0.25	(0.16,0.59)	0.87	(0.58,1.0)	1.8	(1.3,2.4)
			$\alpha = 1.00$			
National	1.40	(0.87,2.1)	0.64	(0.56,0.72)	1.1	(0.97,1.2)
Amazon	0.23	(0.18,0.39)	0.94	(0.77,1.0)	2.1	(1.4,2.5)
Andean	1.70	(1.0,2.8)	0.58	(0.48,0.67)	1.0	(0.94,1.2)
Caribbean	0.82	(0.19,2.00)	0.61	(0.34,0.99)	1.3	(1.0,2.3)
Pacific	0.61	(0.23,1.2)	0.69	(0.51,0.95)	1.2	(0.93,1.9)
Orinoquia	0.26	(0.16,0.53)	0.86	(0.6,1.0)	1.8	(1.3,2.4)

Table 3.5: Parameters estimates from the renewal equation method utilizing the GGM for the early ascending phase of the epidemic (30 days) at the national and regional level.

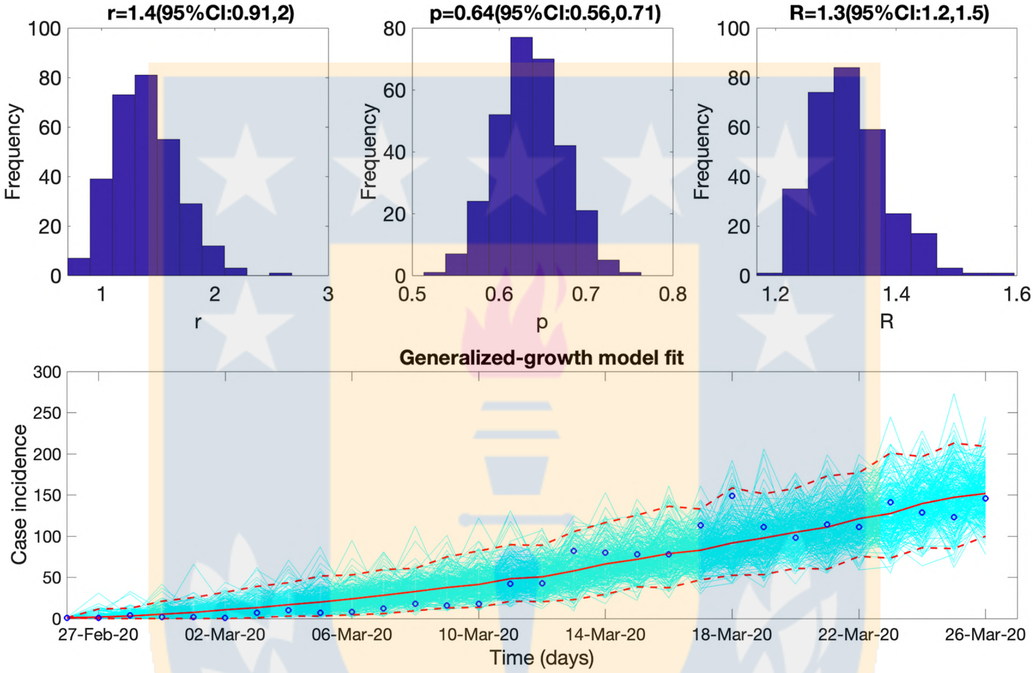


Figure 3.6: Upper panel: Reproduction number for Colombia with 95% CI estimated using the GGM model. The estimated reproduction number of the COVID-19 epidemic in Colombia as of March 27, 2020, is 1.30 (95% CI:$1.20, 1.50$). The growth rate parameter, r, is estimated at 1.40 (95% CI:$0.91, 2.0$) and the deceleration of the growth parameter, p, is estimated at 0.64 (95% CI:$0.56, 0.71$) at $a = 0.15$. Lower panel: The lower panel shows the GGM fit to the case incidence data for the first 30 days from February 27, 2020 to March 27, 2020. The blue circles correspond to the data points; the solid red line indicates the best model fit, and the red dashed lines represent the 95% confidence interval. The cyan lines are the model fits obtained via bootstrapping. The figure is taken from the published article [150].

CHAPTER 4

Sensitivity and identifiability analysis for a model of the propagation and control of COVID-19 in Chile

This chapter shows an extension of an identifiability analysis applied to compartmental models verified in the literature, to be applied to a complex compartmental model inspired by the transmission of COVID-19. This extension incorporates a computational approach and parameter estimation of a variable number of parameters to characterize structural and practical identifiability. Some applications to Chilean regional data are developed.

4.1 Introduction

4.1.1 Scope

The COVID-19 pandemic is currently the main topic of daily news worldwide. We recall that the coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was declared a global pandemic by the World Health Organization (WHO) on March 11, 2020 [28,185]. This highly contagious unprecedented virus has impacted government and public institutions, strained the health care systems, restricted people in their homes, and caused country-wide lockdowns resulting in a global economic crisis [71,91,177]. The impact of COVID-19 at the time of revision of this work (May 3, 2022) amounts to roughly 514 million cases and 6.24 million deaths worldwide of which 2.72 millon have been reported in the Americas.

The morbidity, mortality, and economic indicators associated with the COVID-19 pandemic point to a devastating picture for Latin American countries. High poverty rates, poor leadership, high prevalence of underlying health conditions such as obesity and diabetes, and frail healthcare systems have exacerbated the impact of the novel coronavirus in this region [23,75,86,111]. In Chile, the Ministry of Health reported the first COVID-19 case on March 3, 2020, becoming the fifth country in Latin America after Brazil, Mexico, Ecuador and Argentina to report

COVID-19 cases. Soon after, the Chilean government put in place a number of interventions including the closure of all daycares, schools, and universities (March 16), border controls, telework recommendations (March 18), closure of non-essential businesses (March 19), national night curfews (March 22), as well as targeted lockdowns at the level of municipalities since March 26, 2020. As of May 03, 2022, Chile has accumulated a total of 3563650 cases and 57544 deaths [104]. It is also worth noting that Chile has tested at a higher rate than any other Latin American country [164]. Fortunately, Chile's mass COVID-19 vaccination campaign with the goal of immunizing about 80% of the total population started in February 2021, and meanwhile more than 90% of that target population have been fully vaccinated as of May 03, 2022.

A way to understand and explain these epidemic phenomena is by applying a mathematical model to fit data. Still, within the context of outbreaks of infectious diseases, the information recovered by these epidemic models needs to have guaranteed accuracy or confidence to understand the biological mechanisms and then develop optimal strategies for public health. However, the successful application of such epidemic models depends upon our ability to estimate transmission parameters key, subject to two principal sources of uncertainty: the data's noise and the assumptions built in the model. Ignoring these factors can result in misleading inferences and potentially incorrect public health policy decisions [29, 32, 134].

We inspired by the dynamics of COVID-19 that occurred in Chile in the initial phase of 2020 to explore these factors in parameter estimation. Then, we propose a compartmental model with nonlinear differential equations for the Chilean dynamics, and we use it to analyze the noise from the data and the structure. With this in mind, we developed an identifiability study, employing a computational approach, such as is defined in ref [134]. The identifiability analysis, as explained in ref. [126], determines how well the parameters of a model can be estimated from experimental data, so it is crucial for interpreting and determining confidence in model parameter values. Mainly, the structural identifiability focuses on a mathematical model and the practical identifiability on the data. Hence a model with structurally identifiable parameters may still be nonidentifiable in practice due to a lack of information in real-world data. For that reason, it is essential to understand both identifications to apply an epidemic model at the time [84, 102, 134]. Therefore, we use the compartmental model and the computational approach to investigate the identifications of the parameters. Specifically, we employ synthetic data generated from our model, and we select different sets of parameters to estimate with the help of a defined optimization process with the simulated annealing method. Fitting all data points, we desire to recover the parameters assumed to create synthetic data. If it occurs, it can suggest structural identifiability, which we pretend to validate by adding an uncertainty study with a bootstrapping process [32, 134]. For the case of practical identifiability, we use the same synthetic data but fit only initial phases (e.g., pre-peak). We applied our parameter sets identifiable to fit regional Chilean data to corroborate the conclusions derived from the previous analysis with synthetic data.

4.1.2 Related work

To put the paper into the proper perspective, we first recall that introductions to compartmental epidemiological models are provided in [2, 13, 61, 96, 176, 188]. This class of models goes back to the well-known work by Kermack and McKendrick [85]. Within a compartmental model the total population is subdivided into at least two epidemiological compartments (say, susceptible and infected; but many others can be considered). The rates of progression between the compartments, as well as the incidence and possibly birth and death of individuals, need to be specified, which leads to a system of coupled, and mostly nonlinear ordinary differential equations (ODEs) that describe the progression of the epidemic. In many cases, the compartmental approach in conjunction with suitable assumptions on the rates of progression between the compartments and formulated as a dynamical system provides indications of the basic reproduction number R_0, where this parameter plays the role of a threshold value for the dynamics of the system. Epidemiologically, R_0 gives the number of secondary cases one infectious individual will produce in a population consisting only of susceptible individuals [96, p. 21].

But the conclusions obtained by a compartment model are derived from parameters estimated and allow describing the progression or dynamics to be present in a data set. Then, the identifiability analysis surge as a need to guarantee the good representation given by the parameters estimates. Specifically, the identifiability analysis addresses the question of whether the parameters of a compartment model can be uniquely identified. There are different strategies for the identifiability, for example, for models expressed by linear differential equations [101, 187], nonlinear differential equations [102], or partial differential equations [126], even so, the type of data to fits also plays an essential role, for example, select between incidence or prevalence curves [29, 68, 162]. In this theory are considered two identifiability, one structural and the other practical. The first involves the model and its structure, and the study can be centered on theoretical or computational techniques (e.g., [30, 68, 102, 126]), and the second involves the data (e.g., [84, 162]). The analysis is separate because a model can be structurally identifiable, but the real-world data employed to fit the model results in uncertain parameters.

Consequently, understanding the importance of the identifiability study, we decide to extend the computational approach developed in [134] to assess parameters identifiability in a compartmental model defined by nonlinear ordinary differential equations. Then, the computational approach that combines the optimization process determined by the simulated annealing method and a bootstrap process allows us to access the analysis of the estimated parameter sets. In particular, the structural is assessed by employing synthetic data with parameters known, and the practical, with the same synthetic data, but to fit only initial phases of curve data (e.g., pre-peak moment). In both compute, we wait to recover the parameter set assumed together to a low uncertain and an error small, which traduce in a model with parameters identifiable.

Finally, the identifiability analysis is applied to different parameter sets to estimate. We vary their elements in an increasing number, intending to observe the model's ability to estimate the different sets fitting the data curves generated with assumed parameters. Here we have a sensitivity analysis.

In an application to the Chilean data, we consider the parameter sets identifiable from our computational approach (with synthetic data) and fit some Regional Chile data. We hope to confirm the conclusions from the identifiability study with our application to the Chilean case.

4.2 Methods

4.2.1 Compartmental model

We apply our methodology to a complex compartmental transmission model, whose design is inspired to model by the strategies implemented by the Chilean government to mitigate the COVID-19 emergency in the initial phase of the outbreak. For this purpose, we consider a single population and adopt a simplified version of the model described in ref. [31], combined with the way individuals in quarantine are described in [81]. Individuals within the model are classified as susceptible (S), in home quarantine (Q), latent (E_1), partially infectious but not yet symptomatic (E_2), asymptomatic (A), infectious and will not be tested (I_{n}), infectious and will be tested and isolated (I_{s}), hospitalized/isolated infectious (J), recovered (R), and deceased (D). Constant population size is assumed, i.e.,

$$N := S + Q + E_1 + E_2 + A + I_{\mathrm{n}} + I_{\mathrm{s}} + J + R + D = \mathrm{const}. \tag{4.1}$$

Note that we stipulate one single class of asymptomatic individuals, while in [31] a distinction is made between individuals that are “asymptomatic and will not be tested” and those that are “asymptomatic and will be tested and will be isolated”.

It is assumed that for the home-quarantined individuals, due to severe travel restrictions and rigorous supervision by their local communities, they do not have contact with infected individuals. The parameters p and $1/\lambda$ represent the percentage of individuals in quarantine and the quarantine duration, respectively. Therefore, the class Q has the effect of removing susceptible individuals from the infection dynamics, when $p, \lambda \neq 0$, and there is no quarantine when $p = \lambda = 0$.

Five classes can contribute to new infections, namely E_2, A, I_{n}, I_{s}, and J. If we denote by $r_{X \to Y}$ the rate at which individuals move from class X to class Y, then susceptible individuals move to E_1 at rate

$$r_{S \to E_1} = \frac{\beta}{N}(q_{\mathrm{e}} E_2 + q_{\mathrm{a}} q A + q I_{\mathrm{n}} + q I_{\mathrm{s}} + hJ),$$

where β denotes the overall transmission rate. Individuals in E_1 progress to E_2 at rate κ_1. Individuals from E_2 are partially infectious, with relative transmissibility q_{e}, and progress at rate κ_2, where a proportion ρ_a become asymptomatic and partially infectious (relative transmissibility q_a), and $1-\rho_\mathrm{a}$ become fully infectious. Among the proportion $1-\rho_\mathrm{a}$ that become fully infectious, ρ_s will be tested, while $1-\rho_\mathrm{s}$ will be undetected. Asymptomatic individuals who are not tested and symptomatic individuals wear personal protective equipment (PPE) (such as wearing masks in public, handwashing, etc.) and thus have relative transmissibility q, which quantifies the effectiveness of those protective behaviors. Individuals within classes A and I_n (who are not tested) recover at rate γ_1. Those who are tested (class I_s) will progress to the hospitalized and isolated class at diagnosis rate α. Relative transmission within hospitals and isolated places may occur, but we assume perfect isolation in our analysis. Individuals who are hospitalized and isolated progress to the recovered class at rate γ_2 or to the deceased class at rate δ. Therefore, the model is defined by the following system of non-linear differential equations for a single population, where all variables in capital letters are functions of time t, i.e. it is understood that $S=S(t)$, $E_1=E_1(t)$, etc., and a prime denotes the derivative, that is $S'\equiv \mathrm{d}S/\mathrm{d}t$, etc.

$$\begin{align}
S' &= -\frac{\beta S}{N}(q_\mathrm{e}E_2+q_\mathrm{a}qA+qI_\mathrm{n}+qI_\mathrm{s}+hJ)-pS+\lambda Q, \tag{4.2a}\\
Q' &= pS-\lambda Q, \tag{4.2b}\\
E_1' &= \frac{\beta S}{N}(q_\mathrm{e}E_2+q_\mathrm{a}qA+qI_\mathrm{n}+qI_\mathrm{s}+hJ)-\kappa_1E_1, \tag{4.2c}\\
E_2' &= \kappa_1E_1-\kappa_2E_2, \tag{4.2d}\\
A' &= \kappa_2\rho_\mathrm{a}E_2-\gamma_1A, \tag{4.2e}\\
I_\mathrm{n}' &= \kappa_2(1-\rho_\mathrm{a})(1-\rho_\mathrm{s})E_2-\gamma_1I_\mathrm{n}, \tag{4.2f}\\
I_\mathrm{s}' &= \kappa_2(1-\rho_\mathrm{a})\rho_\mathrm{s}E_2-\alpha I_\mathrm{s}, \tag{4.2g}\\
J' &= \alpha I_\mathrm{s}-(\gamma_2+\delta)J, \tag{4.2h}\\
R' &= \gamma_1(A+I_\mathrm{n})+\gamma_2J, \tag{4.2i}\\
D' &= \delta J, \tag{4.2j}\\
C' &= \alpha I_\mathrm{s}. \tag{4.2k}
\end{align}$$

The auxiliary variable C tracks the cumulative number of diagnosed/reported cases from the start of the outbreak, and C' represents the new diagnosed cases. It is not a state of the system of equations, but symply a class to track the cumulative incidences cases; meaning, individuals from population are not moving to the class C. The initial conditions of these state variables are denoted by $S(0)$, $Q(0)$, $E_1(0)$, $E_2(0)$, $A(0)$, $I_\mathrm{n}(0)$, $I_\mathrm{s}(0)$, $J(0)$, $R(0)$, $D(0)$, and $C(0)$, where $S(0)=N-Q(0)-E_1(0)-E_2(0)-A(0)-I_\mathrm{n}(0)-I_\mathrm{s}(0)-J(0)-R(0)-D(0)$. A schematic of the transmissions is provided in Figure 4.1.

The basic, control, and effective control reproduction numbers R_0, R_0^c and R_e^c are quantities that allow us to measure the epidemic impact of infectious disease in a population as well as

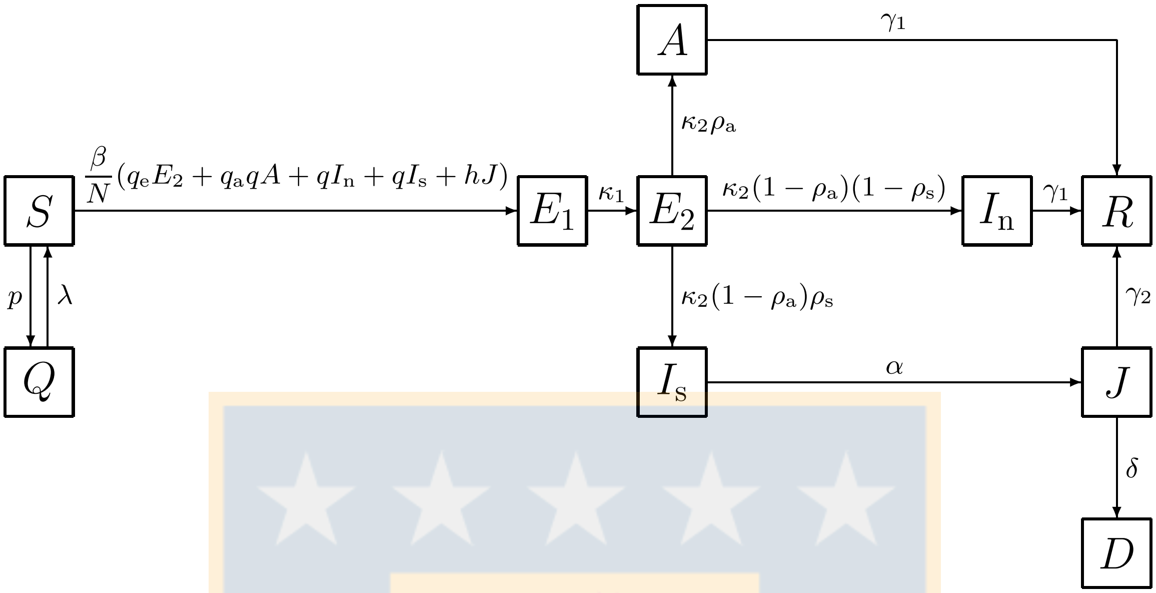


Figure 4.1: Schematic of the transmission of COVID-19 for one population

measuring the effectiveness of control measures. The basic reproduction number R_0 defines the average number of secondary cases generated by primary infectious individuals during the early transmission period when the population is completely susceptible and in the absence of control interventions. The control reproduction number R_0^{c} defines the average number of new cases generated by infectious individuals when the population susceptible is with some control measure or intervention. The effective control reproduction number $R_{\mathrm{e}}^{\mathrm{c}}$ is the reproduction number that varies with time, and it is defined to partially susceptible population, where if $R_{\mathrm{e}}^{\mathrm{c}} > 1$, the disease transmission continues, and if $R_{\mathrm{e}}^{\mathrm{c}} < 1$ the disease transmission declines. Therefore, while the basic reproduction number R_0 and control reproduction number R_0^{c} capture the transmission potential of infectious disease during the early epidemic, the effective control reproduction number $R_{\mathrm{e}}^{\mathrm{c}}$, captures changes in the transmission potential over time.

To calculate these numbers, we employ the well-known next-generation matrix approach [166]. This yields the expression

$$R_0^c = \rho(\boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{V}}^{-1})$$

where ρ denotes the spectral radius and for the present model, the matrices $\boldsymbol{\mathcal{F}}$ and $\boldsymbol{\mathcal{V}}$ are given by

$$\boldsymbol{\mathcal{F}} = \begin{bmatrix} 0 & \beta q_{\mathrm{e}} & \beta q_{\mathrm{a}} q & \beta q & \beta q & \beta h \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\boldsymbol{\mathcal{V}} = \begin{bmatrix} \kappa_1 & 0 & 0 & 0 & 0 & 0 \\ -\kappa_1 & \kappa_2 & 0 & 0 & 0 & 0 \\ 0 & -\kappa_2\rho_{\mathrm{a}} & \gamma_1 & 0 & 0 & 0 \\ 0 & -\kappa_2(1-\rho_{\mathrm{a}})(1-\rho_{\mathrm{s}}) & 0 & \gamma_1 & 0 & 0 \\ 0 & -\kappa_2(1-\rho_{\mathrm{a}})\rho_{\mathrm{s}} & 0 & 0 & \alpha & 0 \\ 0 & 0 & 0 & 0 & -\alpha & \gamma_2+\delta \end{bmatrix}.$$

The result is

$$R_0^{\mathrm{c}} = \beta\left(\frac{q_{\mathrm{e}}}{\kappa_2} + \frac{q_{\mathrm{a}} q \rho_{\mathrm{a}}}{\gamma_1} + \frac{q(1-\rho_{\mathrm{s}})(1-\rho_{\mathrm{a}})}{\gamma_1} + \frac{q\rho_{\mathrm{s}}(1-\rho_{\mathrm{a}})}{\alpha} + \frac{h\rho_{\mathrm{s}}(1-\rho_{\mathrm{a}})}{\gamma_2+\delta}\right). \tag{4.3}$$

Moreover, the corresponding effective reproduction number is defined as

$$R_{\mathrm{e}}^{\mathrm{c}}(t) = R_0^{\mathrm{c}} S(t)/N, \tag{4.4}$$

where $S(t)/N$ is the proportion of susceptible individuals in the population at time t and we recall that the total population N is assumed to be constant (see (4.1)). From the explicit expression (4.3) we can deduce the following contributions of the individual compartments:

$$R^{E_2} = \frac{\beta q_{\mathrm{e}}}{\kappa_2}, \quad R^{A} = \frac{\beta q_{\mathrm{a}} q \rho_{\mathrm{a}}}{\gamma_1}, \quad R^{I_{\mathrm{n}}} = \frac{\beta q(1-\rho_{\mathrm{s}})(1-\rho_{\mathrm{a}})}{\gamma_1},$$

$$R^{I_{\mathrm{s}}} = \frac{\beta q \rho_{\mathrm{s}}(1-\rho_{\mathrm{a}})}{\alpha}, \quad R^{J} = \frac{\beta h \rho_{\mathrm{s}}(1-\rho_{\mathrm{a}})}{\gamma_2+\delta},$$

where

$$R_0^{\mathrm{c}} = R^{E_2} + R^{A} + R^{I_{\mathrm{n}}} + R^{I_{\mathrm{s}}} + R^{J}. \tag{4.5}$$

This parameter R_0^c is a key epidemic parameter to measure the transmissibility of a pathogen; by the expression (4.3) we can see that this parameter is a function of several parameters of our epidemic model (4.2), including transmission rates and infectious periods of the epidemiological states that contribute to new contagious.

In addition, we can observe that the basic reproduction number R_0 corresponds to the control reproduction number R_0^c with absence of interventions, i.e., when $q = 1,\ h = 0,\ \rho_s = 0$, ($q$ is interpreted as the level of no protection, and $(1-q)$ is the level of protection due to the use of PPE), obtaining the following expression,

$$R_0 = \beta\left(\frac{q_{\mathrm{e}}}{\kappa_2} + \frac{q_{\mathrm{a}}\rho_{\mathrm{a}}}{\gamma_1} + \frac{(1-\rho_{\mathrm{a}})}{\gamma_1}\right). \tag{4.6}$$

Parameter	Description	Value	Source
β	transmission rate	1.0676, 1.2435, 1.7794	Assumed
h	relative isolation transmissibility of infected individuals	0	Assumed
q_{e}	relative transmissibility of exposed individuals	0.1	[37]
q_{a}	relative transmissibility of asymptomatic cases	0.4	[37]
q	relative transmissibility of asymptomatic and infected cases that wear PPE	0.4	Assumed
$1/\kappa_1$	length of latent period	2.5	[120, 194]
$1/\kappa_2$	length of infectiousness prior to symptom onset	2.5	[120, 194]
ρ_{a}	proportion of exposed individuals who become asymptomatically infected	0.4	[107, 115]
ρ_{s}	proportion of fully infectious individuals who undergo testing	0.6	Assumed
$1/\alpha$	time from symptom onset to isolation or hospitalization	10	Assumed
$1/\gamma_1$	time from illness onset to recovery	7	[120, 193]
$1/\gamma_2$	time from diagnosis to recovery	5	[154]
δ	death rate within hospitals	0.021	Computed Chilean data
p	proportion of the susceptible population in quarantine declared	0.01, 0.05	Assumed
$1/\lambda$	duration of quarantine	15	Assumed
R_0^c	control reproduction number	3,4,5	Assumed

Table 4.1: Model parameters values selected to generate the simulated data of our study. The parameter R_0^c is fixed, and the expression β is computed for each one using the expression 4.3

4.2.2 Simulated data

Using our mode proposal for COVID-19, we simulate different scenarios fixing their parameters and initial conditions. These curves will help analyze the performance of our model to capture such dynamics. Then a parameter estimation process is necessary together with a bootstrapping approach to examine the identifiability of several of its parameters, as will be explained in the following

For this, we assume the values and references summarized in the Table 4.1 for the parameters, and for the initial conditions the values $J(0) = C(0) = 1$, $Q(0) = D(0) = R(0) = 0$, $I_{\mathrm{n}}(0) = 1$, $I_{\mathrm{s}}(0) = 20$, $E_1(0) = 2J(0)$, $\quad E_2(0) = 4J(0) - E_1(0)$, $A(0) = 4$, and $S(0) = N - E_1(0) - E_2(0) - I_{\mathrm{n}}(0) - I_{\mathrm{s}}(0) - A(0) - J(0) - R(0) - D(0)$, being the population $N = 500000$. All to simulate the following scenarios.

Scenario 1: Initial phase for three R_0^c values without quarantine.

We consider three values for R_0^c, which are generating three values for β, using the expression (4.3), such values together with other parameters are summarized in Table 4.1, and their incidences curves, C' are generated for $t \in [0, 200]$, as we can see in the Figure 4.2. These incidence curves will be our datasets for the identifiability study. Observing these curves, we

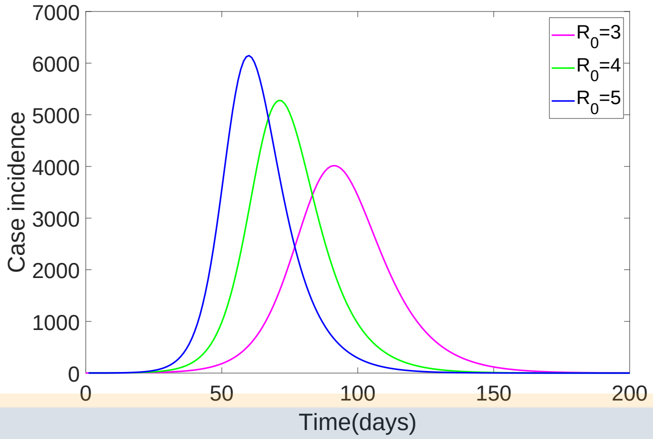


Figure 4.2: Synthetic data to Scenario 1 using the parameters summarized in Table 4.1 where magenta, green and blue lines correspond to the curves generated assumed R_0^c equal to 3, 4, and 5, respectively.

note that to $R_0^c = 3$, the curve effectively grows more slowly than the other curves, being its maximum value lower than the maximums of the curves with $R_0^c = 4$ and $R_0^c = 5$.

Scenario 2: Initial phase for three R_0^c values with quarantine.

For this scenario, we consider the same three values for R_0^c, which produce the same β to scenario 1, because expression (4.3) does not depend of p and λ parameters, then assumed the pairs ($p = 0.01$, $\lambda = 1/15$), and ($p = 0.05$, $\lambda = 1/15$) with the rest of the parameters assumed in Table 4.1, we have the datasets present in Figures 4.3 (A)-(B).

We can observe that curves with $R_0^c = 3$ grow more slowly than the other curves, being its maximum value lower than the maximums of the curves with $R_0^c = 4$ and $R_0^c = 5$. Besides, curves with $p = 0.05$ their maximum values are lower than maximums with $p = 0.01$, which evidences the effect of the quarantine modeled by our model.

We pretend to use these curves defined for each scenario to generate multiple samples from the best-fit model using the bootstrapping approach. But, before we need to define the parameter estimation process to explain the bootstrap approach and finally the identifiability study.

4.2.3 Parameter Estimation

To estimate parameter values of our COVID-19 model, we need to use one optimization process to fit the model to each simulated data. Then, the method used for this purpose is outlined as follows, executed in Matlab (Mathworks, Inc).

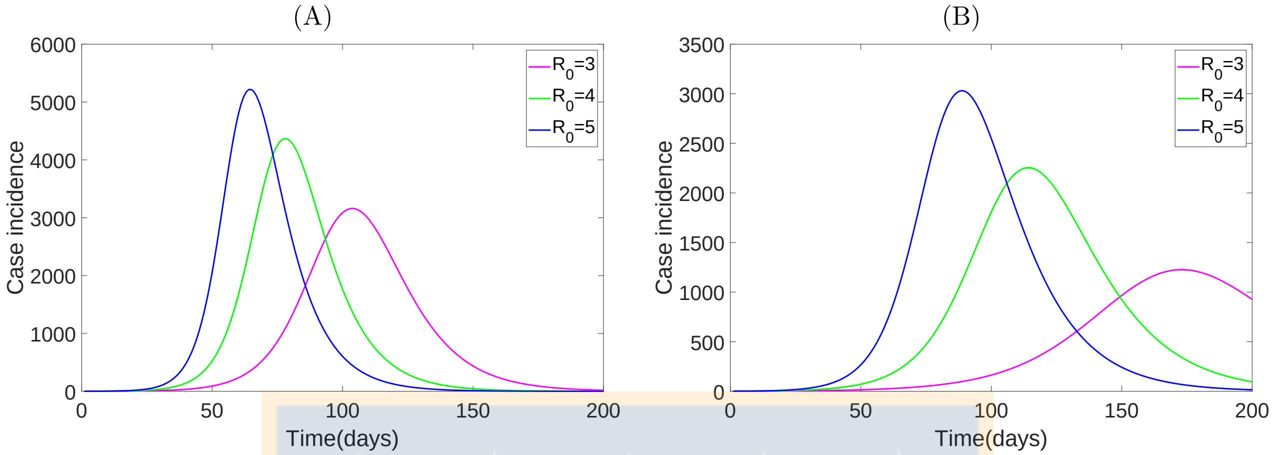


Figure 4.3: Synthetic data to Scenario 2 with $p = 0.01$ (A) and $p = 0.05$ (B) , using the parameters summarize in Table 4.1. The magenta, green and blue lines correspond to the curves generated assumed R_0^c equal to 3, 4, and 5 respectively.

Mixed simulated annealing method with a least-squares approach (SA-LSQ)

To estimate the parameters, we start by defining the parameter set to estimate as Θ, then using `LHS` latin hypercube sampling provided by `lhsdesign()` we generate an initial guess denoted by $\vec{\Theta}$ which the simulated annealing method (SA) is applied to minimize the squared Euclidean distance between the C' curve of our model (4.2) and the data, denoted by data_{t_i}, that corresponds to data in series time with $t \in [0, T]$, with T final date. Therefore, we obtain a best-fit parameter set $\hat{\Theta}$. Specifically, we uses the Matlab function `simulannealbnd()` to minimize the objective function

$$f(\Theta, R) := \frac{1}{R} \sum_{t_i \leq T} \left| C'(t_i, \Theta) - \text{data}_{t_i} \right|^2$$

based on an initial guess $\vec{\Theta}$ for given norming coefficient R. The initial temperature (within the SA approach) is set to 100, and the number of steps is limited to 1000. Further, the parameters optimized are recalled to the unit-hypercube. Finally this process is applied Runs times (a number proportional to the length of the parameter vector Θ) and the best-fit parameter set obtained is verifying with the application of `LeastSquares(`data, $\hat{\Theta}$`)`, using the Matlab function `lsqcurvefit()`.

The step-by-step procedure is summarized in Algorithm 2.

This method is a powerful tool for optimization, in particular for fitting models that describe epidemic phenomena with real data, such as those developed in [16,97,190,192] Additionally; the SA technique was inspired by Metropolis et al. [100], for the selection of the new characterized solutions. Even so, its adaptation allows the application to a frequentist paradigm.

Algorithm 2: Region-wise mixed simulated annealing/least-squares approach.

```
load data
$R \leftarrow 1$
$\vec{\Theta} \leftarrow$ LHS
$\Theta \leftarrow \arg\min_{\Theta \in \vec{\Theta}} f(\Theta, R)$
$R \leftarrow f(\Theta, R)$
res $\leftarrow \infty$
for $i = 1$ to $i <$ Runs do
    $p \leftarrow \mathcal{N}(0, \sigma)$
    $\text{res}_i, \Theta_i \leftarrow$ SimulatedAnnealing(data, $\vec{\Theta}; 2^p R$)
    if $\text{res}_i \cdot 2^p R <$ res then
        res $\leftarrow \text{res}_i \cdot 2^p R$
        $\Theta \leftarrow \Theta_i$
    end if
end for
$\hat{\Theta} \leftarrow$ LeastSquares(data, $\Theta$)
return $\hat{\Theta}$
```

Specifically, to study parameter identifiability, we define different experiments, which use the generated curves in scenarios 1 and 2, and we pretend to estimate or recover from these curves some parameters of interest. Therefore, between the parameters of interest we have set, $\{\alpha, \beta, \rho_s, I_s(0), J(0), p\}$, being the rest permanently fixed according to the values registered in Table 4.1.

For the data generated using scenario 1, we define the following parameter sets to estimate,

1. Parameter β
2. Parameters (α, β)
3. Parameters (β, ρ_s)
4. Parameters (α, β, ρ_s)
5. Parameters $(\beta, I_s(0), J(0))$
6. Parameters $(\alpha, \beta, I_s(0), J(0))$
7. Parameters $(\beta, \rho_s, I_s(0), J(0))$
8. Parameters $(\alpha, \beta, \rho_s, I_s(0), J(0))$

And using the scenario 2, the following parameter set,

1. Parameter β
2. Parameters (α, β)
3. Parameters (β, ρ_s)
4. Parameters (α, β, ρ_s)
5. Parameters (β, p)
6. Parameters (α, β, p)
7. Parameters $(\beta, \alpha, \rho_s, p)$
8. Parameters $(\beta, I_s(0), J(0))$

9. Parameters $(\alpha, \beta, I_s(0), J(0))$
10. Parameters $(\beta, \rho_s, I_s(0), J(0))$
11. Parameters $(\alpha, \beta, \rho_s, I_s(0), J(0))$
12. Parameters $(\beta, I_s(0), J(0), p)$
13. Parameters $(\beta, \alpha, I_s(0), J(0), p)$
14. Parameters $(\alpha, \beta, \rho_s, I_s(0), J(0), p)$

Each combination explored will be mentioned as an experiment. For example, Experiment 101 and Experiment 201-1, where the first case corresponds to data from Scenario 1 to estimate the parameter set 1, and the second case to Scenario 2 with $p = 0.01$ to estimate the respective parameter set 1.

Therefore, for Scenario 1, we have 24 experiments, 3 for each R_0^c assumed and 8 for the number of parameters to estimate. Moreover, for Scenario 2, we have 84 experiments, 3 for each R_0^c, 2 for each p, and 14 for the number of parameters to estimate.

4.2.4 Bootstrapping method

This method uses the parametric bootstrapping approach with Poisson error structure (see [32, 134]) to repeatedly sample observations from the best-fit obtained by the SA-LSQ fit of our COVID-19 model to an time series data. This process can be summarized in the following steps:

1. Obtain the deterministic model solution $C(t, \hat{\Theta})$ or total daily incidence series from the best-fit obtined using the SA-LSQ estimation .
2. Generate S replicates datasets assuming the Poisson error structure. To generate these simulated data , we incorporing the Poisson error structure using the incidence curve $C'(t, \hat{\Theta})$, where for each time t_i we generate a new incidence value using a Poisson random variable with mean $C'(t_i, \hat{\Theta})$. Therefore, this new dataset represents an incidence curve for the system, assuming the time series follows a Poisson distribution centered on the mean at time points t_i.
3. Re-estimate model parameters for each of the S simulated realizations (using the SA-LQS method), which are denoted by $\hat{\Theta}_i$ for $i = 1, ..., S$.
4. Finally, using the set of $\hat{\Theta}_i$ for $i = 1, ..., S$, we construct the confidence intervals for each estimated parameter. Also, for each set of estimated paramters, R_0^c is calculated to obtain a distribution of R_0^c values as well.

4.2.5 Parameter identifiability

The complexity from our mathematical model involves ten epidemiological states or compartments and the auxiliary variable $C(t)$ that indicates the symptomatic cumulative number of diagnosed or reported new cases of the outbreak. In addition to the 10 system states, our model consists of 15 parameters (one-time-dependent and the rest constants), where 5 classes are contributing to a new infection ($E_2, A, I_\mathrm{n}, I_\mathrm{s}, J$), which also contribute to the representation of R_0^c in these classes, see (4.3) and (4.5). Using these expressions, R_0^c is not directly estimated from our model because it is a composite parameter that can be calculated using the individual parameters estimated.

Some model's initial conditions are unknown and maybe need to be estimated, along with the other model parameters. For our case we consider the values $I_s(0)$, and $J(0)$ to estimate together to the parameters α, β, ρ_s and p, then, we have the set from parameters to estimate, as follows,

$$\{\alpha, \beta, \rho_\mathrm{s}, p, I_s(0), J(0)\},$$

being the rest parameters, and initial condition values known or assumed.

To estimate the parameter sets, we fit the model (with other parameters and initial values fixed) to each dataset using the auxiliary variable C' and the curve of the daily cases generated. The parameter estimation procedure employs the methods detailed in Subsection 4.2.3. The initial parameter prediction affects the solution for the model as local minima occur. One way to guarantee the global minimum obtained in the optimization processes, we decided to generate a multi-start (using LHS in Algorithms) repeating these processes for a random amount of initial parameters, whose ranges are defined as follows:

$$\begin{aligned} &0 < \beta < 5, \quad 0.1 < \alpha < 2, \quad 0 < J(0) < C(0), \\ &0 < I_\mathrm{s}(0) < 300, \quad 0 < \rho_\mathrm{s} < 1, \quad \text{and } 0 < p < 1. \end{aligned}$$

We are focusing the parameter identifiability study for our COVID-19 model (4.2) on characterizing the identifiable parameters using uncertainty studies to assess the stability of the parameters estimated. Then, the goal is to evaluate the identifiability using the mean value, the 95% confidence intervals, and the Mean Square Error (MSE).

The step-by-step to study the identifiability is detailed in the next subsection.

Identifiability analysis process

In the study of identifiability, we must follow the following steps that rely on definitions and computations presented in detail in [32] and [134]:

1. Being $\Theta = \{\alpha, \beta, \rho_\mathrm{s}, p, I_s(0), J(0)\}$, the parameter set to estimate and Θ_0 their initial guesses.

2. Estimate Θ fitting our model to the dataset generated, using the SA-LSQ procedure with the values of the initial conditions and parameters given. The symbol $\hat{}$ indicates an

estimated parameter. Then, we have the following set,

$$\hat{\Theta} = \{\hat{\alpha}, \hat{\beta}, \hat{\rho}_{\rm s}, \hat{p}, \hat{I}_s(0), \hat{J}(0)\}.$$

3. Using the bootstrapping process we obtain the 95% confidence intervals and their frequency distributions for each parameter estimated.

4. Compute the mean squared error (MSE) for each parameter estimated. This quantity is calculated as

$$\text{MSE} = \frac{1}{S}\sum_{j=1}^{S}(\Theta - \hat{\Theta}_j)^2,$$

where Θ represents the true parameter value, and $\hat{\Theta}_j$ represents the estimated value of the parameter for the jth bootstrap realization.

5. For each set of estimated parameters, R_0^c is calculated to obtain a distribution of R_0^c values as well.

Following the process described, we say that a model parameter is identifiable if its confidence interval lies in a finite range of values and contains the parameter estimated, preferably if its value has a central position or is close to the mean value. Besides, a small confidence interval indicates that the parameter can be precisely identified, while a broader range could indicate a lack of identifiability. When a parameter can be estimated with low MSE and little confidence interval, the parameter is identifiable from the model. On the other hand, larger confidence intervals or larger MSE values may suggest non-identifiability.

Specifically, to analyze the structural identifiability, we will fit the entire curves generated for each scenario to recover the values from the parameters assumed in each case. Moreover, to investigate the practical identifiability, we will consider different fit times, representing the situation when only existing data of the initial phase (before peak or maximum value).

4.3 Results

4.3.1 Structural analysis

Applied the identifiability methodology, we obtain results for each step, i.e., the first results correspond to the parameter estimation process, which gets the best-fit model solution. Second, we have 250 replicated datasets applying the bootstrap process to re-estimate model parameters. Figures B.3-B.4 show the confidence intervals (represented by vertical lines) and the MSE bar plots for experiment scenario 1, and Figures B.4-B.6 the same graphs-type but for experiments scenario 2.

Observing the results obtained and registered in the figures mentioned, we concentrate our investigation in the Tables 4.2 and 4.3, in which we classify the parameters estimated between identifiable and non-identifiable. We use the green color in the first situation, and it occurs when the parameter estimated has a narrow confidence interval, a mean value-centered and close to real value, and the MSE is small. Furthermore, the second situation is colored red and occurs when at least one identifiability criteria are not satisfied.

From Table 4.2, we conclude that for experiments 1-4, i,e., without initial conditions to fit, we have that all these estimated parameters are identifiable. However, for experiments 5-7, where the initial conditions are estimated, we note that all the parameters except the initial condition $J(0)$ are identifiable, maintaining mainly the identifiability of the parameter of interest R_0^c. Only experiment 8 shows non-identifiability situations for all the parameters, except for the dataset generated assuming $R_0^c = 5$. We can suspect that the identifiability improves when the epidemic overgrows, as in this last case. Nevertheless, we will continue to explore this part later. Another observation is that the estimated R_0^c can be robust to variation or bias in the other estimated parameters.

In the same way from Table 4.3, we conclude that the parameters estimated in experiments 1, 2, 3, and 5 are identifiable, besides the parameters except for the initial condition $J(0)$ in the experiments 8, 9, 10, and 12 are also identifiable. For experiments 4, 7, and 9, the parameter R_0^c is identifiable even though none of its other parameters are identifiable. Particularly in experiment 7, all the rest parameters are non-identifiable. Other situations occur; for example, in experiment 6, all parameters estimated except using the data generated with $R_0^c = 3, p = 0.05$ are identifiable. For experiment 13, all parameters estimated using data generated using $R_0^c = 4$ and 5 are identifiable. For experiment 14, the results are not good, but the R_0^c estimated using data generated with $p = 0.01$ are identifiable. A similar situation occurs with R_0^c in experiment 11, where it is identifiable only fitting the data generated with $R_0^c = 5$, and $p = 0.1$.

4.3.2 Practical analysis

The results obtained for this part are in the Appendix B, where colored tables were created using the same definitions and criteria as for Tables 4.2 and 4.3. Then in Tables B.3-B.4 from Appendix B, we can analyze the identifiability. For this study, we investigate the performance of our model to capture the initial phases of an epidemic disease, establishing three fit times, that is, 20, 30, and 40 days.

Observing Table B.3, we can see that in experiment 1, the parameters estimated are identifiable. In experiments 3 and 5 with fit times 30 and 40, their parameters R_0^c are identifiable, even though not all parameters estimated are identifiable. In contrast to this result, in experiment 2, fitting a time of 20 days, the parameters that define R_0^c are identifiable, but the R_0^c computed is not. Notably, for this experiment, the confidence intervals and the MSE are plotted in Figure B.7, which we can see that these intervals are not very large but lack precision, mainly when

Par. Est.	R_0	Experiment							
		1	2	3	4	5	6	7	8
α	3	⬛	🟩	⬛	🟩	⬛	🟩	⬛	🟥
	4	⬛	🟩	⬛	🟩	⬛	🟩	⬛	🟥
	5	⬛	🟩	⬛	🟩	⬛	🟩	⬛	🟩
β	3	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟥
	4	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟥
	5	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
ρ_s	3	⬛	⬛	🟩	🟩	⬛	⬛	🟩	🟥
	4	⬛	⬛	🟩	🟩	⬛	⬛	🟩	🟥
	5	⬛	⬛	🟩	🟩	⬛	⬛	🟩	🟩
$I_s(0)$	3	⬛	⬛	⬛	⬛	🟩	🟩	🟩	🟥
	4	⬛	⬛	⬛	⬛	🟩	🟩	🟩	🟥
	5	⬛	⬛	⬛	⬛	🟩	🟩	🟩	🟩
$J(0)$	3	⬛	⬛	⬛	⬛	🟥	🟥	🟥	🟥
	4	⬛	⬛	⬛	⬛	🟥	🟥	🟥	🟥
	5	⬛	⬛	⬛	⬛	🟥	🟥	🟥	🟥
R_0	3	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟥
	4	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟥
	5	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩

Table 4.2: Summarize of results obtained for Scenario 1. The green box represents that the parameter estimated satisfies the parameter identifiability conditions, i.e., this has small interval confidence with mean value-centered and a low MSE. The parameter that does not meet the identifiability criteria is colored red boxes.

are used 20 days in the fit, and the epidemic growth is given by $R_0^c = 3$, and 4. Apparently, for this and other experiments, the parameter R_0^c tends to be identifiable when fit times are greater than 20 days, and the growth assumed R_0^c is lower or equal to 4.

Parameters ρ_s, and the initial condition $I_s(0)$ and $J(0)$ for all experiments from this scenario 1 are not identifiable. Suggesting a lack of identifiability, mainly from parameter ρ_s when we have fewer data to fit.

Table B.4 also notes that the parameters estimated for experiment 1 are identifiable. It is different from experiments 12, 13, and 14, where all their parameters are non-identifiable. For this scenario 2, we can observe that for most experiments with a time of 20 days to fitting (except experiments 1 and 9 with $R_0^c = 3$), the R_0^c is non-identifiable. Moreover, all parameters estimated are also non-identifiable, especially for these cases, excluding experiment 2. Suggesting perhaps the necessity of more fit data to guarantee the identifiability, such as we can see occur more frequently in experiments with more fit times.

Another situation to comment on here is that for most experiments where the quarantine parameter p is estimated, the results mainly for R_0^c result non-identifiable, excluding only experiment 5 for fit time 40, and R_0^c assumed as 4 and 5, and experiment 6 for fit time 40, and $R_0^c = 3$, and 4. Besides, for some quarantine parameters estimated from experiments 5 and 6 with fit time 40, the result is identifiable, mostly when the fit data are generated assuming $p = 0.01$. We can think that when more population is quarantined in the initial phase, the lack of identifiability increases.

4.4 Application to regional Chilean data

We decided to apply our methodology to the data collected for some Chilean regions. From the structural and practical results obtained for our model, we will apply the study to two cases, one to scan the initial phase (first 30 days) and another for an intermediate phase (first 120 days), considering the data before the change of quarantine criteria applied by the Chilean government. To explain the region selection and the phenomena of the COVID-19 in Chile, we include the following Subsection.

4.4.1 The early transmission of COVID-19 in Chile

We wish to apply the COVID-19 model of Section 4.2.1 to some administrative regions of Chile (see Figure 4.4), where there are regions without quarantine periods, others with more than one quarantine period, a situation that the Chilean government named “dynamical quarantines”. This policy was in effect until July 27 2020. In addition, this measure was applied only to municipalities with a greater incidence of positive cases for COVID-19. For this reason, we consider a decoupled analysis to select the regions to analyze until July 27 2020, assuming the day of the first positive case as the first day for the timeline of each region and July 27 2020 as the last day. Since the onset of COVID-19 in Chile, the Ministry of Health has been reporting new cases of infection and death daily for each region and new cases of recovered persons at the national level. The official information (see [103, 105] for official details) also includes the daily numbers of polymerase chain reaction (PCR) tests applied, as well as the daily numbers of critical patients and occupancy of intensive care units (ICUs). A timeline of strategies and measures implemented to contain the pandemic is summarized in Table 4.4. Several times the Chilean Ministry of Health changed the case and death definitions. Table 4.5 lists the most relevant of these changes along with their dates. Considering the difficulties associated with the changes in case definitions, we decided to include data reported for symptomatic and asymptomatic in the COVID-19 model. Total data correspond to the addition of symptomatic and asymptomatic cases reported because an asymptomatic (tested) case may become symptomatic later.

Par. Est.	p	R_0	Experiment													
			1	2	3	4	5	6	7	8	9	10	11	12	13	14
α	0.01	3	⬜	🟩	⬜	🟩	⬜	🟩	🟥	⬜	🟩	⬜	🟥	⬜	🟥	🟥
		4	⬜	🟩	⬜	🟩	⬜	🟩	🟥	⬜	🟩	⬜	🟥	⬜	🟩	🟥
		5	⬜	🟩	⬜	🟩	⬜	🟩	🟥	⬜	🟩	⬜	🟩	⬜	🟩	🟥
	0.05	3	⬜	🟩	⬜	🟥	⬜	🟥	🟥	⬜	🟥	⬜	🟥	⬜	🟥	🟥
		4	⬜	🟩	⬜	🟥	⬜	🟩	🟥	⬜	🟩	⬜	🟥	⬜	🟩	🟥
		5	⬜	🟩	⬜	🟩	⬜	🟩	🟥	⬜	🟩	⬜	🟥	⬜	🟩	🟥
β	0.01	3	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟩	🟥	🟩	🟥	🟥
		4	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟩	🟥	🟩	🟩	🟥
		5	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟩	🟥	🟩	🟩	🟥
	0.05	3	🟩	🟩	🟩	🟥	🟩	🟩	🟥	🟩	🟥	🟩	🟥	🟩	🟥	🟥
		4	🟩	🟩	🟩	🟥	🟩	🟩	🟥	🟩	🟩	🟩	🟥	🟩	🟩	🟥
		5	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟩	🟥	🟩	🟩	🟥
ρ_s	0.01	3	⬜	⬜	🟩	🟩	⬜	⬜	🟥	⬜	⬜	🟩	🟥	⬜	⬜	🟥
		4	⬜	⬜	🟩	🟩	⬜	⬜	🟥	⬜	⬜	🟩	🟥	⬜	⬜	🟥
		5	⬜	⬜	🟩	🟩	⬜	⬜	🟥	⬜	⬜	🟩	🟥	⬜	⬜	🟥
	0.05	3	⬜	⬜	🟩	🟥	⬜	⬜	🟥	⬜	⬜	🟩	🟥	⬜	⬜	🟥
		4	⬜	⬜	🟩	🟥	⬜	⬜	🟥	⬜	⬜	🟩	🟩	⬜	⬜	🟥
		5	⬜	⬜	🟩	🟩	⬜	⬜	🟥	⬜	⬜	🟩	🟥	⬜	⬜	🟥
$I_s(0)$	0.01	3	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟩	🟩	🟩	🟥	🟩	🟥	🟥
		4	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟩	🟩	🟩	🟥	🟩	🟩	🟥
		5	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟩	🟩	🟩	🟥	🟩	🟩	🟥
	0.05	3	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟩	🟩	🟩	🟥	🟩	🟥	🟥
		4	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟩	🟩	🟩	🟥	🟩	🟩	🟥
		5	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟩	🟩	🟩	🟥	🟩	🟩	🟥
$J(0)$	0.01	3	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		4	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		5	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟥	🟥	🟥	🟥	🟥	🟥	🟥
	0.05	3	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		4	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		5	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟥	🟥	🟥	🟥	🟥	🟥	🟥
p	0.01	3	⬜	⬜	⬜	⬜	🟩	🟩	🟩	⬜	⬜	⬜	⬜	🟩	🟥	🟥
		4	⬜	⬜	⬜	⬜	🟩	🟩	🟥	⬜	⬜	⬜	⬜	🟩	🟩	🟥
		5	⬜	⬜	⬜	⬜	🟩	🟩	🟥	⬜	⬜	⬜	⬜	🟩	🟩	🟥
	0.05	3	⬜	⬜	⬜	⬜	🟩	🟥	🟥	⬜	⬜	⬜	⬜	🟩	🟥	🟥
		4	⬜	⬜	⬜	⬜	🟩	🟩	🟥	⬜	⬜	⬜	⬜	🟩	🟩	🟥
		5	⬜	⬜	⬜	⬜	🟩	🟩	🟥	⬜	⬜	⬜	⬜	🟩	🟩	🟥
R_0	0.01	3	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟥	🟩
		4	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟩
		5	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
	0.05	3	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟩	🟩	🟥	🟩	🟥	🟥
		4	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟥
		5	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟥

Table 4.3: Summarize of results obtained for Scenario 2. The green box represents that the parameter estimated satisfies the parameter identifiability conditions, i.e., this has small interval confidence with mean value-centered and a low MSE. The parameter that does not meet the identifiability criteria is colored red boxes.

Date	Measure
16 Mar	Closing of schools and universities
18 Mar	Declaration of national emergency and border closure
	Massive tests using the PCR test in each region
19 Mar	Store closings except for pharmacies, banks, and supermarkets
21 Mar	Closing of entertainment centers
22 Mar	Declaration of national curfew between 10:00 PM and 5:00 AM
26 Mar	Start of lockdown and quarantine in different municipalities
8 Apr	Mandatory use of face masks in public spaces
25 Jul	The Chilean government declared a gradual re-opening plan (“Paso a Paso”) with different levels of permitted activities at each step. Municipalities may move forward or backward in steps, depending on local conditions

Table 4.4: The Chilean government applied the following public health interventions during the first wave in 2020 to contain the COVID-19 epidemic.

Date	New criterion
29 Apr	Incorporation status “asymptomatic,” i.e., people without symptoms but with PCR positive.
1 Jun	Deaths without a diagnosis but with suspected symptoms of COVID-19 or indeterminate PCR are added to official daily reports.
2 Jun	Report of PCR tests applied for each region is suspended (for approx. a week).
3 Jun	Report of recovered cases is suspended.
7 Jun	653 deaths added as possible cases of COVID-19 (patients with indeterminate PCR or with symptoms similar to those caused by COVID-19), and 96 deaths of people diagnosed do not have a precise regional distribution in the reports, due to a change in the region of residence
17 Jun	The status “without notifying” is created in the daily reports, summing 31422 cases of people with PCR + who were not reported in the system on diagnosis (reports delayed).

Table 4.5: Changes in criteria of reported data

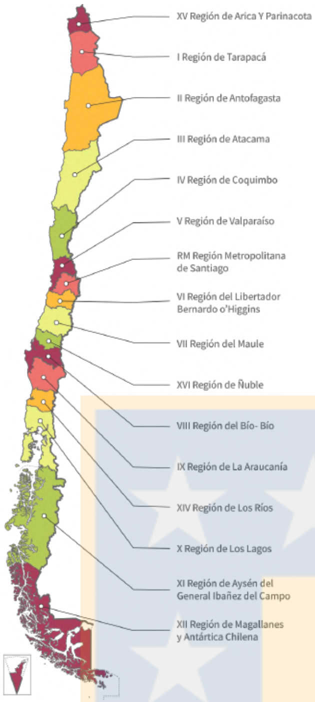
4.4.2 Experiments applied to regional data

Understanding and analyzing the type of information and data published by the Chilean government, we find regions with more than one quarantine period. Now, comparing it with the conclusions obtained by our identifiability study, where we have results from experiments with a population without quarantine periods and only one quarantine. Then, with this in mind, we select administrative regions that satisfy some of these conduct,

1. Have one unique quarantine period for the first 30 days or 120 days.

2. The first 30 days or 120 days do not have quarantine periods.

We desire that the data has a growing number of reported cases in both situations, i.e., do not have consecutive zeros cases (robust data).


XV Región de Arica Y Parinacota
I Región de Tarapacá
II Región de Antofagasta
III Región de Atacama
IV Región de Coquimbo
V Región de Valparaíso
RM Región Metropolitana de Santiago
VI Región del Libertador Bernardo o'Higgins
VII Región del Maule
XVI Región de Ñuble
VIII Región del Bío- Bío
IX Región de La Araucanía
XIV Región de Los Ríos
X Región de Los Lagos
XI Región de Aysén del General Ibañez del Campo
XII Región de Magallanes y Antártica Chilena

Number	Region	population
15	Arica	226.068
1	Tarapacá	330.558
2	Antofagasta	607.534
3	Atacama	286.168
4	Coquimbo	757.586
5	Valparaíso	1.815.902
6	O'Higgins	914.555
7	Maule	1.044.950
16	Ñuble	480.609
8	Biobío	1.556.805
9	Araucanía	957.224
14	Los Ríos	384.837
10	Los Lagos	828.708
11	Aysén	103.158
12	Magallanes	166.533
13	Metropolitana	7.112.808
total		17.574.003

Figure 4.4: The 16 administrative regions of Chile [54] and their population according to the 2017 census [58]. The Roman numbers are the official administrative numbers of the geographic regions. The greater Santiago area is the Metropolitan region (RM) and counted as region 13. Note that numbering is not strictly ordered from north to south; regions 14 to 16 have been created by dividing existing regions. The total population at the end of 2020 is estimated at 19.1 million.

Overall, the experimental results, using simulated data to analyze the identifiability of the parameters (structural and practical), the best results are achieved when are estimated the sets (β, R_0^c), and $(\beta, J(0), I_s(0), R_0^c)$. However, we will try to observe the quarantine parameter, estimating the sets (β, p, R_0^c), and $(\beta, p, J(0), I_s(0), R_0^c)$, for the case that the region has a single quarantine period, even knowing that these parameters are not always identifiable from the experimental results in the initial phase. However, we will try to observe the quarantine parameter, estimating the sets (β, p, R_0^c), and $(\beta, p, J(0), I_s(0), R_0^c)$, for the case that the region has a single quarantine period, even knowing that these parameters are not always identifiable from the experimental results. Particularly in the initial phase, more details in the Table 4.6.

Subsequently, we filter the regional data for two periods, one for the first 30 days and the other for 120 days, and we select some of those that meet the conditions numbered before. We have established the following regions for our application.

- Regions 1 (Tarapacá) and 8 (Bío-Bío), without quarantine periods in the initial phase and one quarantine in the final phase.
- Regions 3 (Atacama) and 4 (Coquimbo) were without quarantine during the 120 days.

Parameter set	Without Quarantine		With Quarantine	
	Structural	Practical	Structural	Practical
(β, R_0^c)	Identifiable	Identifiable	Identifiable	Identifiable
$(\beta, J(0), I_s(0), R_0^c)$	Identifiable	Identifiable $T_{fit} > 20$ days	Identifiable	Identifiable $T_{fit} > 20$ days
(β, p, R_0^c)			Identifiable	Identifiable $T_{fit} = 40$ $R_0^c = 5$ and $p = 0.01$
$(\beta, p, J(0), I_s(0), R_0^c)$			Identifiable	Nonidentifiable

Table 4.6: Summary of the identifiable parameters concluded from the structural and practical identifiability study used in the application to Chilean data.

- Region 10 (Los Lagos) had one quarantine in the initial phase.

Then, with this regional data, we applied our identifiability study to two events,

1. Regions without quarantine, estimating the following sets

   Experiment 1: (β, R_0^c)
   Experiment 2: $(\beta, J(0), I_s(0), R_0^c)$

   Depending on the phase analyzed, we selected the time of 30 or 120 days to fit.

2. Regions with quarantine, estimating the following sets

   Experiment 1: (β, p, R_0^c)
   Experiment 2: $(\beta, p, J(0), I_s(0), R_0^c)$

   We are preserving the same consideration for the time fit.

Left plot in figures 4.5, and 4.6 displays the regional data distributions, including some dates of quarantine measures and important holidays.

Establishing the experiments and the Chilean data for the analysis process, we need to fix the values of the parameters which are not estimated. We consider the same parameters taken for the analysis with synthetic data, only including the values for $1/\alpha = 5$ and $\rho_2 = 0.5$, like averages. These assumptions are in Table B.5 in Appendix B.

Quarantine parameters λ and p are assumed as zero when the experiment does not involve quarantine periods. Alternatively, the parameter p is estimated depending on the region, and the parameter $1/\lambda$ takes values about the quarantine duration applied to it (See Appendix B, Table B.6).

Initial contiditions satisfy the considerations assumed in the identifiabilty study, i.e.,

$$R(0) = Q(0) = D(0) = 0,\ E_1(0) = 2J(0),\ E_2(0) = 4J(0) - E_1(0),\ I_n(0) = \min J(0), 1,$$

$$S(0) = N - E_1(0) - E_2(0) - I_n(0) - I_s(0) - A(0) - J(0),$$

where for the case when $J(0)$, $I_s(0)$ are estimated, these comply $0 < J(0) < C(0)$, and $0 < I_s(0) < 300$. But, when these are assumed, we consider their values as $J(0) = 1$ and $I_s(0) = 10$.

Exp.	T_{fit}	region	β	p	$I_s(0)$	$J(0)$	R_0^c comp	presnorm
				without quarantine				
1	30	1	1.10459				2.30859	1.40E+02
	30	8	1.55171				3.24307	3.17E+03
	120	3	0.69428				1.45104	5.28E+03
	120	4	0.73344				1.53288	8.88E+04
2	30	1	1.08156		11.23401	1.00000	2.26045	1.40E+02
	30	8	1.12597		52.89334	1.00000	2.35327	2.37E+03
	120	3	0.68112		14.52313	0.99994	1.42353	5.23E+03
	120	4	0.61346		148.95176	2.00000	1.28214	5.63E+04
				with quarantine				
1	30	10	2.80707	0.08965			5.86678	9.20E+02
	120	1	2.54383	0.05503			5.31660	1.21E+05
	120	8	5.00000	0.54340			10.45000	1.44E+05
2	30	10	3.70074	0.12446	5.43419	0.99998	7.73455	9.11E+02
	120	1	1.50472	0.03014	53.68752	0.61428	3.14486	1.05E+05
	120	8	3.53811	0.41259	62.03312	0.98066	7.39466	1.10E+05

Table 4.7: Parameter estimation results for the application with Chilean data.

4.4.3 Application results

Finally, applying our methodology to Chilean regions data selected for the experiments defined, we obtained the fits plotted in Figure 4.5-4.6, the colored Table 4.8 which we constructed following the same procedure as the last colored tables for the structural and practical analysis. Where are observed the 95% confidence intervals obtained by the 250 replicate of Bootstrapping process and the MSE (See Tables B.8-B.9, and Figures B.10-B.16). Studying the colored table, we note that in most cases, the parameters β and R_0^c (computed) are identifiable, except Region 10, which has a quarantine period in the first 30 days. Maybe having one quarantine with few points to fit, our model does not have success, and the estimations generated are non-identifiable. As expected, given the results summary in the table 4.6. Region 8 has a short quarantine period in the middle of the final phase, and fitting it region to 120 days from data has difficulty estimating the parameter sets identifiable. On the other hand, Region 1 with a lengthy quarantine period in the final phase has trouble estimating the parameters in experiment 2. Perhaps the initial conditions include more noise in the estimation. In short of these two last regions, we can see that the duration of quarantine and the phase analyzed can be decisive when estimating parameters; these examples suggest that we need to explore more combinations of quarantine periods in our COVID-19 model. In addition to this, the parameters estimated from experiments applied to Regions 3 and 4 are identifiable, as expected, given the structural and practical identifiability demonstrated for our model using our methodology. In the same way, we verify the identifiability obtained for experiments 1 and 2 applied to Region 8, and Region 1 has only parameters identifiable when we use experiment 2. However, their confidence intervals are not distant when we apply experiment 1. Maybe the noise contributed by initial conditions is more prevalent in Region 1 than in Region 8.

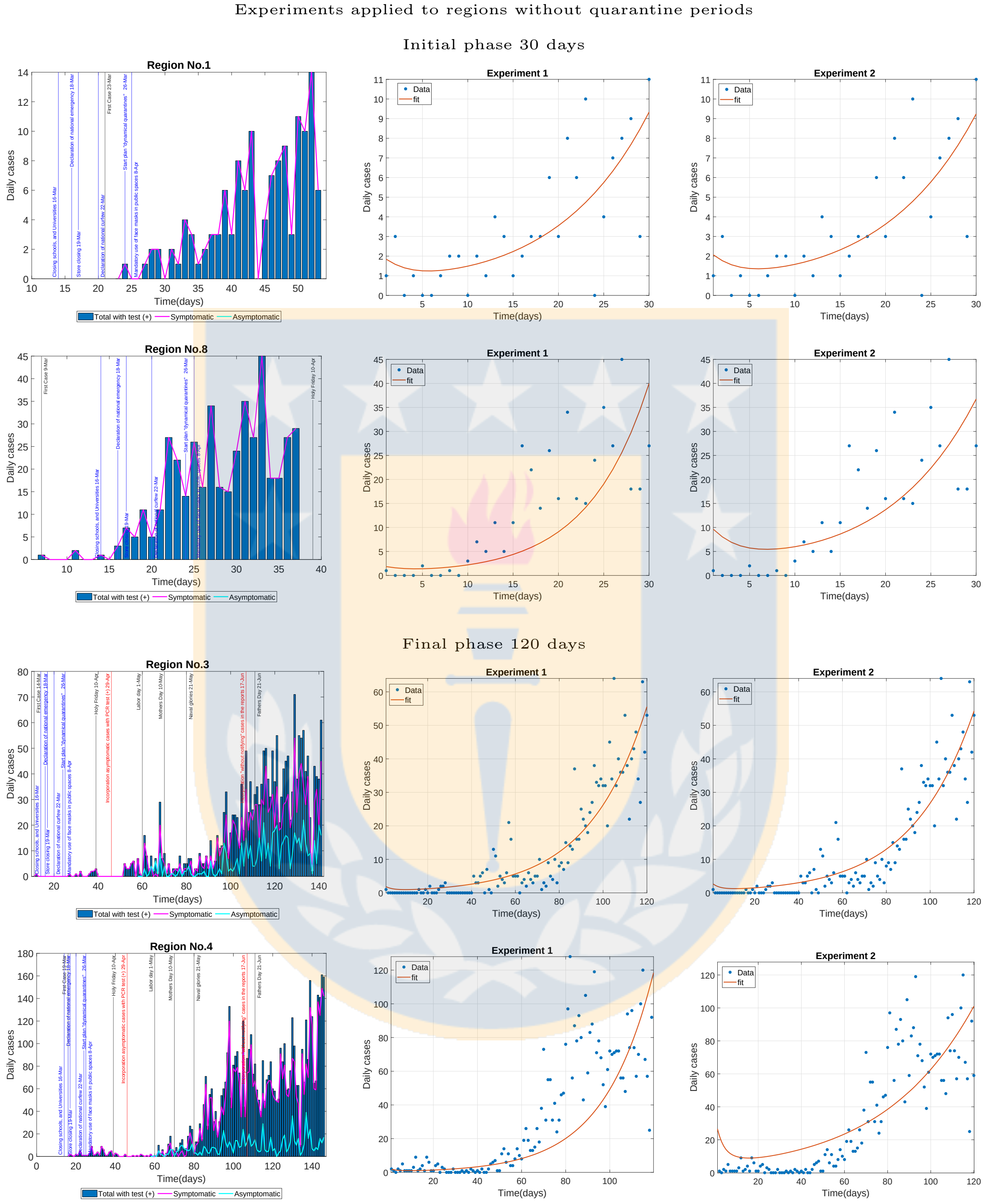


Figure 4.5: The left plot has blue bars representing the daily cases reported in the region considered to apply experiments 1 and 2. The magenta and cyan lines show a proportion between symptomatic and asymptomatic patients. The plots in the middle and right represent the regional data (blue points) and the best-fit (red line) obtained by each experiment (estimated parameter sets are summarized in Table 4.7).

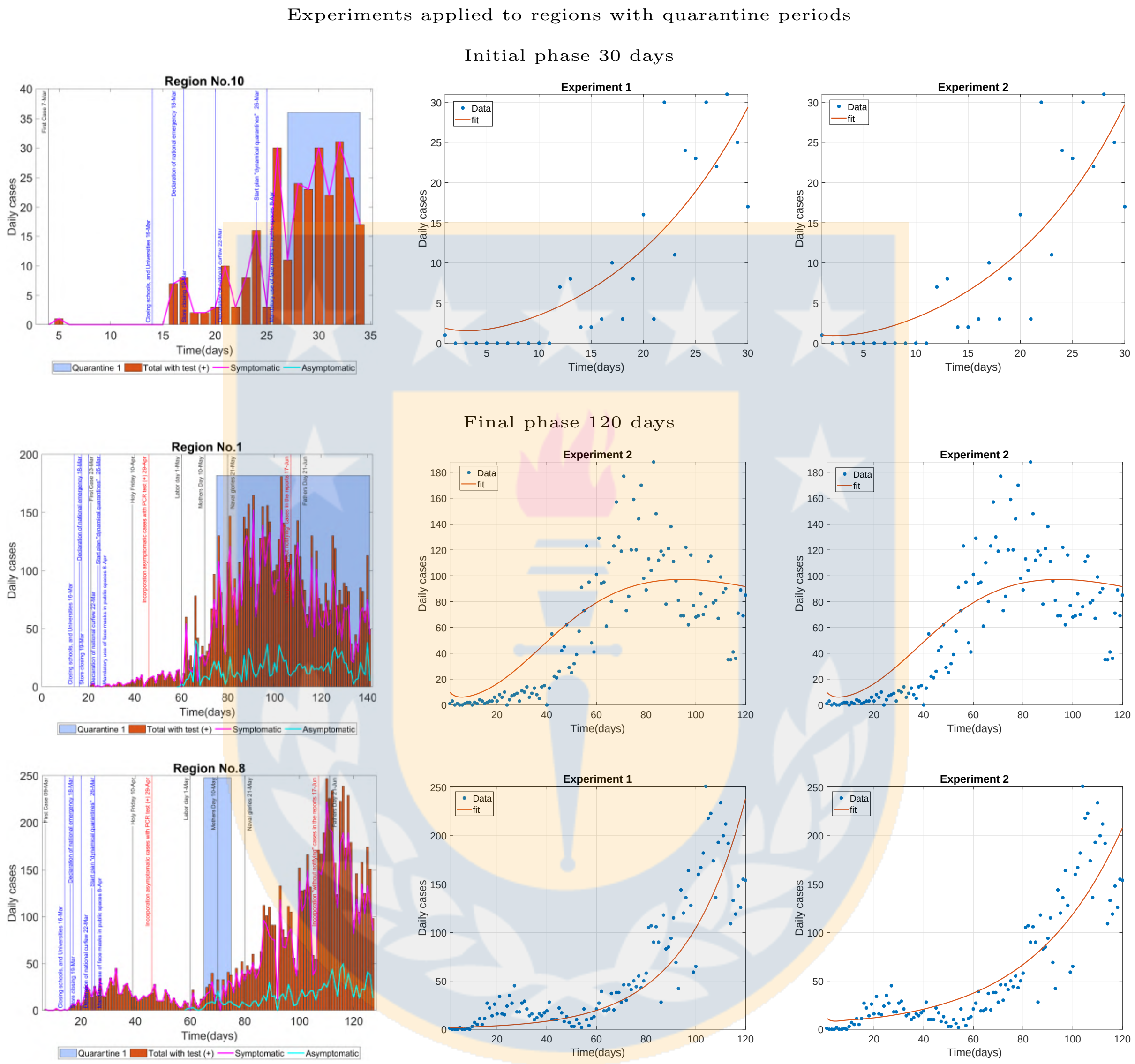


Figure 4.6: The left plot has blue bars representing the daily cases reported in the region considered to apply experiments 1 and 2. The magenta and cyan lines show a proportion between symptomatic and asymptomatic patients. The blue shaded areas indicate the quarantine period declared. The plots in the middle and right represent the regional data (blue points) and the best-fit (red line) obtained by each experiment (estimated parameter sets are summarized in Table 4.7).

Par. Est.	Region	Without quarantine				With quarantine			
		30		120		30		120	
		1	2	1	2	1	2	1	2
β	1	🟩	🟥	▨	▨	▨	▨	🟩	🟥
	3	▨	▨	🟩	🟩	▨	▨	▨	▨
	4	▨	▨	🟩	🟩	▨	▨	▨	▨
	8	🟩	🟩	▨	▨	▨	▨	🟥	🟥
	10	▨	▨	▨	▨	🟥	🟥	▨	▨
p	1	▨	▨	▨	▨	▨	▨	🟩	🟥
	3	▨	▨	▨	▨	▨	▨	▨	▨
	4	▨	▨	▨	▨	▨	▨	▨	▨
	8	▨	▨	▨	▨	▨	▨	🟥	🟥
	10	▨	▨	▨	▨	🟥	🟥	▨	▨
$J(0)$	1	▨	🟥	▨	▨	▨	▨	▨	🟥
	3	▨	▨	▨	🟥	▨	▨	▨	▨
	4	▨	▨	▨	🟥	▨	▨	▨	▨
	8	▨	🟥	▨	▨	▨	▨	▨	🟥
	10	▨	▨	▨	▨	▨	🟥	▨	▨
$I_s(0)$	1	▨	🟥	▨	▨	▨	▨	▨	🟥
	3	▨	▨	▨	🟥	▨	▨	▨	▨
	4	▨	▨	▨	🟥	▨	▨	▨	▨
	8	▨	🟥	▨	▨	▨	▨	▨	🟥
	10	▨	▨	▨	▨	▨	🟥	▨	▨
R_0^c	1	🟩	🟥	▨	▨	▨	▨	🟩	🟥
	3	▨	▨	🟩	🟩	▨	▨	▨	▨
	4	▨	▨	🟩	🟩	▨	▨	▨	▨
	8	🟩	🟩	▨	▨	▨	▨	🟥	🟥
	10	▨	▨	▨	▨	🟥	🟥	▨	▨

Table 4.8: Parameter identifiability verification. The parameter sets estimated in the experiments fitting regional Chilean data are classified between identifiable and non-identifiable. The green box represents that the parameter estimated satisfies the parameter identifiability conditions, i.e., the confidence intervals (obtained through the Bootstrapping process) are narrow and their mean values have a central position in the gap; furthermore, its MSE is small. The parameter that does not meet the identifiability criteria is colored red boxes and the cases not studied are painted gray.

CHAPTER 5

Conclusions and future works

Conclusions

Here we present a summary with the main contributions and conclusions of the thesis.

- In Chapter 1 our systematic comparison of a number of epidemic outbreaks using phenomenological growth models indicates that the GLM outperformed the other models in describing the great majority of the epidemic trajectories. In a few cases (such as Cases 3, 4, 23, and 28) the GGoM outperformed the other models. These findings indicate that the parameter p plays a much more significant role in shaping the dynamic trajectories supported by the GLM compared to the GoM since we observed that the errors of the GoM and GGoM models stay fairly close to each other and the contribution of the adjustment of p remains subtle in some cases. In fact, a closer examination of the parameter estimates derived from both models GoM and GGoM indicates that parameter p is close to 1 in these models, which explains the similarity in the fits derived from these models. So the GGoM model could be reduced to GoM without much impact on the model fit. This is in sharp contrast to what is happening with the logistic models where both the LM and GLM models only yield similar fits for three epidemics. Future research could be directed at determining which of the models equipped with generalized growth are easier to calibrate than the other, considering the initial or final parts of the dynamics and with the aim to improve predictions.

  Referring to the parameter estimation procedure and the need to provide an initial solution to the optimization numerical methods, we have found that Matlab functions and the steps defined in the section 1.3.3, are sufficient for the present study, in agreement with the experience made in [32, 170]. However, since there is a limited range for some of the parameters (as is the case of parameter p, but not of the others) it might be interesting in future work to use metaheuristic procedures to the parameter estimation that possibly guarantee in an appropriate form that the parameters found are indeed optimal globally. As is mentioned in [130], such procedures include simulated annealing (see, e.g., [128,132,133]), variable neighborhood search (VNS) [132,133], and the so-called firefly algorithm [8].

While we compared phenomenological growth models based on their ability to describe empirical trajectories of real epidemics, our methodology could be extended to assess the “distance" between models in terms of the range of dynamics supported by model A that can also be supported by model B and vice versa. For instance, based on our empirical findings we hypothesize that the distance between the LM and GLM models is larger compared to the distance between the GoM and GGoM models. Importantly such distance could be derived for any pair of models regardless of model complexity. Future work could explore this research direction by analyzing a larger set of dynamic models including phenomenological and mechanistic models.

- In Chapter 2, first of all, let us recall that the purpose of this work is not primarily a fit of determined phenomenological growth models to specific data but to introduce a general methodology of applying statistics to medicine and biology. Nevertheless, we may briefly comment the specific outcome for the five models studied in this paper before coming to possible extensions.

  Overall, we can say that in light of the results of the application of the methodology to different types of growth curves, the GLM produces curves that are closer to the (simulated) data than other models, and the curves produced by the GGoM are most distant to the other models. Besides, the results indicate that introducing the parameter p within the GLM and RM significantly improves the adjustment compared with the original logistic model (LM), while most results obtained with the GGoM lead to parameters $p \approx 1$ in most fits, that is, the GGoM is essentially reduced to the GoM with parameters $\Theta = (r, b)$ with $p = 1$. To further highlight the value of the GLM, we mention that this model does not only better approximate the dynamics of data obtained by simulation with other models but as Section 2.4 illustrates, also captures better real data due the advantageous contribution of the growth scaling parameter p. This fact is also demonstrated in our previous work [22]. A possible “mechanistic” explanation of the superiority of the GLM could be related to the different degrees of influence of the parameters. For example, within the GLM we have $C(t) \to K$ as $t \to \infty$, so we simply need to adjust K to specify a final size of the epidemic while varying r and p does not affect this property. In contrast, as follows from (2.10) (see also [22]), for the GGoM with $0 < p < 1$ we have

$$C(t) \to \bigl((1-p)(r/b) + C(0)^{1-p}\bigr)^{1/(1-p)} \quad \text{as } t \to \infty,$$

  which means that the *final* size of the epidemic depends on a number of parameters, in particular the exponent p that is supposed to characterize *early* growth, and on the initial size of the epidemic $C(0)$, which is usually a small number that can hardly be determined with certainty. Probably the fact that within the GLM the early and late stages of the epidemic are dominantly influenced by different parameters, namely p and r; and K, respectively, provides an advantage for reliable parameter identification.

Our interest in the range $0 < p < 1$ for the GLM and GGoM comes from the wish to characterize *sub*-exponential initial growth, as is motivated in [42, 45, 170]. However, this same parameter p can have another nature in the RM, where there are studies with $p > 1$, for example, the papers [10, 161], where we recall that in [22] it was stated that the parameter p within the RM does not serve as an adjustable parameter to capture sub-exponential initial growth. Rather, by its position within the RM the parameter p could allow the shape of upper part of the cumulative curve to be independent of the shape of the lower part, i.e., measures the extent of deviation from the S-shaped dynamics of the classical logistic growth model. Besides, as the parameter p tends to zero, the RM curve tends towards the Gompertz growth curve in the sense $\mathrm{d}C/\mathrm{d}t = rC(t)\ln(K/C(t))$ (see our discussion of the autonomous form of the Gompertz differential equation in Sect. 2.2.1). There are other studies on different forms to generalize PGMs, as [161] which shows for case of logistic growth, different to our idea of generalized growth model with $rC(t)^p$, where p is a scaling parameter. Therefore future work will study the EDDs distances between other generalized PGMs. Then if we consider the range $p > 1$ for the RM, we can see that this model captures the dynamics of influenza data better than the GLM, as is evidenced in Figure 2.13 and Table 2.12.

We emphasize that our restriction to just five PGMs (namely the LM, GLM, GoM, GGoM, and RM) does in no way represent a limitation inherent to the present approach. In fact, it is not the intention of the present work to provide an exhaustive survey of PGMs in epidemiology but to introduce a methodology to compare PGMs within a small selection with each other. In this sense, other models could be examined as well with the same methodology. For instance, one could consider the four-parameter so-called generalized Richards model (GRM; see [32, 36, 119]) given by

$$f(t, C; \Theta) = rC^p\bigl(1 - (C/K)^q\bigr), \quad \Theta = (r, p, q, K), \quad r, q, K > 0, \quad 0 < p \leq 1$$

that combines the generalizations of the GLM (2.2) and the RM (2.5).

Finally, we remark that the methodology of the present work could also be applied to other applications where describing growth by phenomenological models is of interest. As an example, we mentioned in Section 2.1.2 the growth of tumours. In fact, there there is a wealth of alternative phenomenological growth models designed for that application, and to which the present methodology could be applied in future work. We refer to [60, 63, 147] for overviews, and as one specific example the so-called Gomp-ex law (proposed in [181]; see [63]) that for the autonomous form (2.7) can be specified as

$$\varphi(C; \Theta) = (t, C; \Theta) = \begin{cases} C(a - b\ln C_{\text{crit}}) & \text{if } 0 < C < C_{\text{crit}}, \\ C(a - b\ln C) & \text{if } C \geq C_{\text{crit}}, \end{cases} \qquad \Theta = (a, b, C_{\text{crit}}),$$

where the Gompertz law (under suitable choices of the constants a and b) comes into effect only for sufficiently large populations (i.e., whose size is larger or equal to a given critical size C_{crit}), but below C_{crit} growth is exponential [63].

- Chapter 3 characterizes the transmission dynamics of the COVID-19 pandemic in Colombia by fitting mathematical models to national and regional data. Besides, the forecasts are also included. Our results indicate that the sub-epidemic model is the most accurate in terms of calibration and forecasting performances. More importantly, the regional and national level GLM and RM forecasts point towards a continuous declining trend in the epidemic trajectory compared to the sub-epidemic model that can reproduce the sustained growth pattern particularly distinguishable for the national, Caribbean Andean, and the Pacific region. Overall, the transmission dynamics show sustained disease transmission during the early phase of the COVID-19 pandemic exhibiting sub-exponential growth dynamics at the regional and national levels. As the epidemic progressed, fluctuations in R_t we observed, with the most recent estimates that $R_t < 1.00$ indicating disease containment.

  Appropriate short-term forecasts at the national and regional levels can help guide the intensity and magnitude of public health interventions required to contain the epidemic. The short-term forecasts from the GLM and RM indicate a sustained decline in the overall case counts like the forecast produced by the sub-epidemic wave model for the Amazon and Orinoquía region. However, the sub-epidemic model predicts the stabilization in case incidence for the national, Andean, Caribbean, and Pacific region. On the contrary, the mortality curve forecast predicted by the sub-epidemic wave model shows an increase in death. The different projections obtained should be interpreted with caution, given the instability in the reporting patterns and reporting delays. Our analysis shows that the sub-epidemic wave model performs better than the GLM and RM in short-term forecasts based on the performance metrics (Tables 3.4-B.2). This same situation occurs in the short term forecasting of the COVID-19 pandemic applied to México in [149] The sub-epidemic wave model is also a better fit to the epidemic trajectories during the calibration period compared to the other two models (Tables 3.3-B.1).

  The early transmission dynamics of SARS-CoV-2 exhibit similarity at the national and regional levels. The COVID-19 pandemic in Colombia exhibited sub-exponential growth dynamics ($0 < p < 1$) during the early ascending phase of the outbreak at the national and regional levels. Results are consistent with the sub-exponential growth patterns observed in other Latin American countries, including México [98] and Chile [153] which also implemented mask mandates and social distancing interventions along with restricting mobility during the early growth phase of the pandemic. Simultaneously, the estimates of early transmission potential (R_t) indicate sustained disease transmission in Colombia at the national and regional levels with $R_t > 1$. These estimates suggest that although containment strategies were implemented during the first thirty days to mitigate the impact of the pandemic (Figure 3.2), additional interventions should be prioritized, such as obligatory social distancing and intensified case surveillance. The results of our analysis are compatible with the estimates of early reproduction numbers retrieved from other countries, including Peru, Chile, Brazil, and Mexico, which followed similar COVID-19 outbreaks around the same period [70, 109, 149, 153].

This study has some limitations. Such as, the last 11-day case counts are excluded where we utilize the case counts based on the onset dates in this study because delays in case reporting, testing rates, and factors related to the surveillance systems can influence our epidemic projections. Secondly, we relied on the daily updates of cases in the official surveillance system of Colombia, which can sometimes underreport. Third, the PGMs applied in this study do not explicitly account for behavioral changes. Thus the results such as the predicted decline or stability in the epidemic trajectory should interpret with caution. Lastly, the unpredictable social component of the epidemic on the ground was also a limiting factor for the study. When the forecasts were generated, we did not know the ground truth epidemic pattern.

The forecasts need to be interpreted with caution given the spatial heterogeneity in transmission rates and dynamic implementation and lifting of the social distancing measures. The PGMs employed in this study to forecast and estimate reproduction numbers are valuable for providing rapid predictions of the epidemics in complex scenarios that can be used in real-time because these models do not require specific disease transmission processes to account for the interventions.

- In Chapter 4, we present and apply a computational approach developed in ref [134] for simple epidemic models. However, we propose adapting it to a complex compartmental model inspired by the first wave of the COVID-19 outbreak in Chile, where quarantines were declared involving different Chilean regions in early phases. Specifically, we propose a methodology to analyze the structural and practical parameter identifiability in a COVID-19 model. We explore the parameter uncertainty computationally through a parametric bootstrapping approach to achieve that goal. In that exploration, we involve synthetic data generated with the same COVID-19 model to measure the capability of our model to recover the parameters assumed to satisfy the criteria for identifiability structural and practical. Those criteria involve observing the 95% confidence intervals and the MSE (Mean square error), constructed from empirical distributions resulting from Bootstrapping process. In this first attempt, we selected some experiments that would allow us to validate the structural and practical parameter identifiability. We consider different variables, for example, the number of parameters to fit, the time to fit, and the quarantine periods. Then, in our methodology, we construct colored tables (e.g., 4.2) to visualize these qualities. Overall in the results obtained, we observe that parameter β, and in consequence, R_0^c are identifiable (when the rest parameters are known). But, if one more parameter is included, the uncertainty in some cases increases, making the parameter set non-identifiable. Maybe for dependences between it. The following best parameter sets in terms of structural identifiability are the sets without initial conditions to estimate, especially in the case without quarantines involved. Wherein the cases with quarantine are more sensitive as a percentage of the population in quarantine increases and the epidemic growth is $R_0 < 4$, evidence of a lack of identifiability in the model when the quarantine dynamics are considered. Then, I would take quarantine periods studied in the models with caution. Expanding the analysis to practical identifiability in the initial phases, we find a significant loss of parameter identifiability. It compared with the structural results. Verifying maybe the conclusion shown in ref [139] indicates the non-identifiability in pre-peak moments. But, we have an interest in analyzing the initial period to recover the control reproduction number. We consider applying our methodology in different epidemic periods to evaluate our model parameters a process essential. In our experiments, we confirm that the β parameter is recovered, and in the cases when the initial conditions are also estimated, the identifiability is a guarantee for curves with a growth from $t > 20$ days. Show another exciting characteristic, the dependence between the fit times and the growth rates, besides quarantine periods, which also affect the identifiability, but maintain the β parameter identifiable. Finally, collecting the conclusions from the experiment with synthetic data, we select the parameter sets (β), and $(\beta, J(0), I_s(0))$ for periods without quarantine, and (β, p), and $(\beta, p, J(0), I_s(0))$ for periods with quarantine, to be fitting with Chilean data for some regions. We find that the model parameters are not identifiable in the Chilean regions with a quarantine period in the initial phase. On the other hand, the regions without quarantine periods have better results, where between more data points are fitted, better estimations. For future work, we see the necessity of selecting the fit times for each Chilean region because each one has a particular quarantine declared; besides, the quality data also is variable. Then we can compare the fit times exposed here [74] and compare the results. Currently, the values to β parameter for Region 8 in the initial phase are closed.

Future Work

In general terms we can indicate that the probable scenarios for future research are:

- Expand the comparative study of PGMs to other models not considered, which may be helpful in mathematical epidemiology or another line of application.

- Mathematically structure the distance between PGMs, which allows better characterization of such models for adjusting infectious diseases.

- Likewise, from what has been seen so far for the study of COVID-19 in Chile, it is necessary to consider quarantine periods in a discrete timeline to recover the reproduction number more reliably.

- Include a sensitivity study for the estimated parameters fitting the compartment model, which allows us to understand their dependencies.

- Deepen the study of parameter identification in compartmental models to establish a generalized methodology for the computational analysis. We could invertigate new processes to improve the Bootstrapping execution times and new algorithms for estimating parameters, such as the Markov chain Monte Carlo technique (MCMC) or the recent Hamiltonian Monte Carlo.

CHAPTER 6

Conclusiones Generales y Trabajos Futuros

Conclusiones

A continuación, se presenta un resumen con los principales aportes y conclusciones generadas en esta tesis.

- En el Capítulo 1 nuestra comparación sistemática de un número de brotes epidémicos, usando modelos de crecimiento fenomenológicos indica que el GLM describiendo la gran mayoría de trayectorias epidémicas se comporta mejor que los otros modelo. En algunos casos (como los Casos 3, 4, 23, and 28) el GGoM superó a los otros modelos. Estos hallazgos indican que el parámetro p juega un papel mucho más importante en la configuración de las trayectorias dinámicas respaldadas por el GLM en comparación con el GoM, ya que, observamos que los errores de los modelos GoM y GGoM permanecen bastante cerca uno del otro y la contribución del ajuste de p sigue siendo sutil en algunos casos. De hecho, un examen más detallado de las estimaciones de parámetros derivadas de ambos modelos, GoM y GGoM, indica que el parámetro p está cerca de 1 en estos modelos, lo que explica la similitud en los ajustes derivados de estos modelos. Entonces, el modelo GGoM podría reducirse a GoM sin mucho impacto en el ajuste del modelo. Esto contrasta marcadamente con lo que sucede con los modelos logísticos, donde tanto el modelo LM como el GLM solo arrojan ajustes similares para tres epidemias. Futuras investigaciones podrían estar dirigidas a determinar cuáles de los modelos equipados con crecimiento generalizado son más fáciles de calibrar que otros, considerando las partes inicial o final de la dinámica y con el fin de mejorar las predicciones.

  En referencia al procedimiento de estimación de parámetros y la necesidad de dar una solución inicial a los métodos numéricos de optimización, hemos encontrado que las funciones de Matlab y los pasos definidos en la sección1.3.3, son suficientes para el presente estudio, de acuerdo con la experiencia realizada en [32, 170]. Sin embargo, dado que hay un rango limitado para algunos de los parámetros (como es el caso de parámetro p, pero no de los otros) podría ser interesante en trabajos futuros usar procedimientos metaheurísticos para la estimación de parámetros que posiblemente garanticen en forma adecuada que los parámetros encontrados son efectivamente óptimos globalmente. Como se menciona en [130], dichos procedimientos incluyen recocido simulado (ver, por ejemplo, [128,132,133]), búsqueda de vecindad variable (VNS) [132,133], y el llamado algoritmo firefly [8].

Si bien comparamos modelos de crecimiento fenomenológico en función de su capacidad para describir trayectorias empíricas de epidemias reales, nuestra metodología podría extenderse para evaluar la “distancia” entre modelos en términos del rango de dinámica respaldado por el modelo A que también puede ser respaldado por el modelo B y viceversa. Por ejemplo, con base en nuestros hallazgos empíricos, planteamos la hipótesis de que la distancia entre los modelos LM y GLM es mayor en comparación con la distancia entre los modelos GoM y GGoM. Es importante destacar que dicha distancia podría derivarse para cualquier par de modelos, independientemente de complejidad del modelo. Un trabajo futuro podría ser explorar esta dirección de investigación mediante el análisis de un conjunto más amplio de modelos dinámicos, incluidos los modelos fenomenológicos y mecanicistas.

- En el Capítulo 2 En primer lugar, recordemos que el propósito de este trabajo no es principalmente un ajuste de determinados modelos de crecimiento fenomenológico a datos específicos, sino introducir una metodología general de aplicación de la estadística a la medicina y la biología. Sin embargo, podemos comentar brevemente el resultado específico de los cinco modelos estudiados en este trabajo antes de llegar a posibles extensiones.

  En general, podemos decir que a la luz de los resultados de la aplicación de la metodología a diferentes tipos de curvas de crecimiento, el GLM produce curvas que están más cerca de los datos (simulados) que otros modelos, y las curvas producidas por el GGoM son más distanted a las otros modelos. Además, los resultados indican que introduciendo el parámetro p dentro del GLM y RM mejoran significativamente el ajuste en comparación con el modelo logístico original (LM), mientras que la mayoría de los resultados obtenidos con el GGoM conducen a parámetros $p \approx 1$ en la mayoría de los ajustes, es decir, el GGoM se reduce esencialmente al GoM con parámetros $\Theta = (r, b)$ con $p = 1$. Para resaltar aún más el valor del GLM, mencionamos que este modelo no solo se aproxima mejor a la dinámica de datos obtenidos por simulación con otros modelos pero, como ilustra la Sección 2.4, también captura mejores datos reales debido a la contribución ventajosa del parámetro de escala de crecimiento p. Este hecho también se demuestra en nuestro trabajo anterior [22]. Una posible explicación "mecanicista" de la la superioridad del GLM podría estar relacionada con los diferentes grados de influencia de los parámetros. Por ejemplo, dentro del GLM tenemos $C(t) \to K$ como $t \to \infty$, por lo que simplemente necesitamos ajustar K para especificar el tamaño final de la epidemia mientras que variar r y p no afecta esta propiedad. En contraste, como sigue de (2.10) (ver también [22]), para el GGoM con $0 < p < 1$ tenemos

$$C(t) \to \big((1-p)(r/b) + C(0)^{1-p}\big)^{1/(1-p)} \quad \text{as } t \to \infty,$$

lo que significa que el tamaño *final* de la epidemia depende de una serie de parámetros, en particular del exponente p que se supone caracteriza el crecimiento *temprano*, y en el tamaño inicial de la epidemia $C(0)$, que suele ser un número pequeño que difícilmente se puede determinar con certeza. Probablemente el hecho de que dentro del GLM las primeras y últimas etapas de la epidemia está predominantemente influenciada por diferentes parámetros, a saber, p y r; y K, respectivamente, proporciona una ventaja para una identificación fiable de los parámetros.

Nuestro interés en el rango $0 < p < 1$ para GLM y GGoM proviene del deseo de caracterizar el crecimiento inicial *sub*-exponencial, como se motiva en [42, 45, 170]. Sin embargo, este mismo parámetro p puede tener otra naturaleza en la RM, donde hay estudios con $p > 1$, por ejemplo, los trabajos [10, 161], donde recordamos que en [22] se indicó que el parámetro p dentro del RM no sirve como un parámetro ajustable para capturar el crecimiento inicial subexponencial. Más bien, por su posición dentro del RM, el parámetro p podría permitir que la forma de la parte superior de la curva acumulativa sea independiente de la forma de la parte inferior, es decir, mide el grado de desviación de la dinámica en forma de S de el modelo de crecimiento logístico clásico. Además, como el parámetro p tiende a cero, la curva RM tiende hacia la curva de crecimiento de Gompertz en el sentido $\mathrm{d}C/\mathrm{d}t = rC(t)\ln(K/C(t))$ (ver nuestra discusión de la forma autónoma de la ecuación diferencial de Gompertz en la Secc. 2.2.1). Hay otros estudios sobre diferentes formas de generalizar PGMs, como [161] que muestran una idea para el crecimiento logístico, diferente a nuestra idea de modelo de crecimiento generalizado, con $rC(t)^p$, donde p es un parámetro de escalamiento. Por lo tanto, un trabajo futuro podría ser estudiar las distancias EDD entre otros PGM generalizados. Entonces si consideramos el rango $p > 1$ para el RM, podemos ver que este modelo captura la dinámica de los datos de influenza mejor que el GLM, como es evidenciado en la figura 2.13 y Tabla 2.12.

Hacemos hincapié en que nuestra restricción a solo cinco PGMs (a saber, LM, GLM, GoM, GGoM y RM) no representa de ninguna manera una limitación inherente al presente enfoque. De hecho, no es la intención del presente trabajo proporcionar un estudio exhaustivo de los PGMs en epidemiología, sino introducir una metodología para comparar PGMs dentro de una pequeña selección entre sí. En este sentido, otros modelos podrían ser examinados también con la misma metodología. Por ejemplo, se podría considerar el llamado modelo de Richards generalizado (GRM; ver [32, 36, 119]) de cuatro parámetros dado por

$$f(t, C; \Theta) = rC^p\big(1 - (C/K)^q\big), \quad \Theta = (r, p, q, K), \quad r, q, K > 0, \quad 0 < p \le 1$$

que combina las generalizaciones del GLM (2.2) y el RM (2.5).

Finalmente, remarcar que la metodología del presente trabajo también podría aplicarse a otras aplicaciones, donde es de interés describir el crecimiento mediante modelos fenomenológicos. Como ejemplo, mencionamos en la Sección 2.1.2 el crecimiento de tumores, de hecho, hay una gran cantidad de modelos alternativos de crecimiento fenomenológico diseñado para esta aplicación, y en la cual la metodología actual podría aplicarse en trabajos futuros. Nos referimos a [60, 63, 147] para obtener información general, y como ejemplo específico de la llamada ley Gomp-ex (propuesta en [181]; ver [63]) que para la forma autónoma (2.7) puede especificarse como

$$\varphi(C;\Theta) = (t, C;\Theta) = \begin{cases} C(a - b\ln C_{\text{crit}}) & \text{if } 0 < C < C_{\text{crit}}, \\ C(a - b\ln C) & \text{if } C \geq C_{\text{crit}}, \end{cases} \qquad \Theta = (a, b, C_{\text{crit}}),$$

donde la ley de Gompertz (bajo elecciones adecuadas de las constantes a y b) entra en vigor solo para poblaciones grandes (es decir, cuyo tamaño es mayor o igual a un tamaño crítico dado C_{crit}), pero por debajo de C_{crit} el crecimiento es [63] exponencial.

- En el Capítulo 3 se caracteriza la dinámica de transmisión de la pandemia de COVID-19 en Colombia ajustando modelos matemáticos a datos nacionales y regionales. Además, también se incluyen las previsiones. Nuestros resultados indican que el modelo subepidémico es el más preciso en términos de desempeño de calibración y pronóstico. Más importante aún, los pronósticos GLM y RM a nivel regional y nacional apuntan hacia una tendencia descendente continua en la trayectoria epidémica en comparación con el modelo subepidémico que puede reproducir el patrón de crecimiento sostenido particularmente distinguible para la región nacional, el Caribe Andino y el Pacífico. En general, la dinámica de transmisión muestra una transmisión sostenida de la enfermedad durante la fase inicial de la pandemia de COVID-19, mostrando una dinámica de crecimiento subexponencial a nivel regional y nacional. A medida que avanzaba la epidemia, observamos fluctuaciones en R_t, y las estimaciones más recientes de $R_t < 1,00$ indican contención de la enfermedad.

  Los pronósticos a corto plazo apropiados a nivel nacional y regional pueden ayudar a orientar la intensidad y la magnitud de las intervenciones de salud pública necesarias para contener la epidemia. Los pronósticos a corto plazo del GLM y RM indican una disminución sostenida en el conteo general de casos como el pronóstico producido por el modelo de onda subepidémica para la Amazonía y la Orinoquía. Sin embargo, el modelo subepidémico predice la estabilización de la incidencia de casos para la región nacional, Andina, Caribe y Pacífico. Por el contrario, la curva de mortalidad prevista por el modelo de onda subepidémica muestra un aumento de la muerte. Las diferentes proyecciones obtenidas deben interpretarse con cautela, dada la inestabilidad en los patrones de notificación y los retrasos en la notificación. Nuestro análisis muestra que el modelo de onda subepidémica funciona mejor que el GLM y el RM en los pronósticos a corto plazo basados en las métricas de rendimiento (Tablas 3.4-B.2). Esta misma situación ocurre en el pronóstico de corto plazo de la pandemia de COVID-19 aplicado a México en [149]

El modelo de onda subepidémica también se ajusta mejor a las trayectorias epidémicas durante el período de calibración en comparación con los otros dos modelos (Tablas 3.3-B.1).

La dinámica de transmisión temprana del SARS-CoV-2 exhibe similitudes a nivel nacional y regional. La pandemia de COVID-19 en Colombia exhibió una dinámica de crecimiento subexponencial ($0 < p < 1$) durante la fase ascendente temprana del brote a nivel nacional y regional. Los resultados son consistentes con los patrones de crecimiento subexponencial observados en otros países de América Latina, incluidos México [98] y Chile [153], que también implementaron mandatos de máscara e intervenciones de distanciamiento social junto con la restricción de la movilidad durante la fase inicial de crecimiento de la pandemia Simultáneamente, las estimaciones del potencial de transmisión temprana (R_t) indican una transmisión sostenida de la enfermedad en Colombia a nivel nacional y regional con $R_t > 1$. Estas estimaciones sugieren que aunque se implementaron estrategias de contención durante los primeros treinta días para mitigar el impacto de la pandemia (Figura 3.2), se deben priorizar intervenciones adicionales, como el distanciamiento social obligatorio y la intensificación de la vigilancia de casos. Los resultados de nuestro análisis son compatibles con las estimaciones de los números de reproducción temprana recuperados de otros países, incluidos Perú, Chile, Brasil y México, que siguieron brotes similares de COVID-19 en el mismo período [70, 109, 149, 153].

Este estudio tiene algunas limitaciones. Por ejemplo, se excluyen los recuentos de casos de los últimos 11 días cuando utilizamos los recuentos de casos en función de las fechas de inicio en este estudio porque los retrasos en la notificación de casos, las tasas de prueba y los factores relacionados con los sistemas de vigilancia pueden influir en nuestras proyecciones epidémicas. En segundo lugar, nos basamos en las actualizaciones diarias de casos en el sistema de vigilancia oficial de Colombia, que en ocasiones pueden subregistrarse. En tercer lugar, los PGM aplicados en este estudio no tienen en cuenta explícitamente los cambios de comportamiento. Por lo tanto, los resultados como la disminución prevista o la estabilidad en la trayectoria de la epidemia deben interpretarse con cautela. Por último, el componente social impredecible de la epidemia sobre el terreno también fue un factor limitante para el estudio. Cuando se generaron los pronósticos, no conocíamos el patrón epidémico real.

Los pronósticos deben interpretarse con cautela dada la heterogeneidad espacial en las tasas de transmisión y la implementación dinámica y el levantamiento de las medidas de distanciamiento social. Los PGM empleados en este estudio para pronosticar y estimar los números de reproducción son valiosos para proporcionar predicciones rápidas de las epidemias en escenarios complejos que se pueden usar en tiempo real porque estos modelos no requieren procesos específicos de transmisión de enfermedades para dar cuenta de las intervenciones.

- En el capítulo 4, presentamos y aplicamos un enfoque computacional desarrollado en ref. [134] para modelos epidémicos simples. Sin embargo, proponemos adaptarlo a un modelo compartimental complejo inspirado en la primera ola del brote de COVID-19 en Chile, donde se declararon cuarentenas que involucraron a diferentes regiones chilenas en fases tempranas.

Específicamente, proponemos una metodología para analizar la identificabilidad estructural y práctica de parámetros en un modelo COVID-19. Entonces para lograr ese objetivo, exploramos la incertidumbre de los parámetros computacionalmente a través de un enfoque de arranque paramétrico. En esa exploración, involucramos datos sintéticos generados con el mismo modelo COVID-19 para medir la capacidad de nuestro modelo para recuperar los parámetros asumidos satisfaciendo los criterios de identificabilidad estructural y práctica. Dichos criterios implican observar los intervalos de confianza de 95% y el MSE (Mean square error), construidos a partir de lss distribuciones empíricas resultantes del proceso Bootstrapping. En este primer intento, seleccionamos algunos experimentos que nos permitieran validar la identificabilidad estructural y práctica de los parámetros. Consideramos diferentes variables, por ejemplo, la cantidad de parámetros a ajustar, el tiempo de ajuste y los períodos de cuarentena. Luego, en nuestra metodología, construimos tablas de colores (e.g., 4.2) para visualizar estas cualidades. En general, en los resultados obtenidos, observamos que el parámetro β y, en consecuencia, R_0^c son identificables (cuando se conocen los demás parámetros). Pero, si se incluye un parámetro más, la incertidumbre en algunos casos aumenta, haciendo que el conjunto de parámetros no sea identificable. Quizá por dependencias entre estos. Los siguientes mejores conjuntos de parámetros en términos de identificabilidad estructural son los conjuntos sin estimar las condiciones iniciales, especialmente en el caso sin cuarentenas involucradas. Donde los casos con cuarentena son más sensibles a medida que aumenta el porcentaje de la población en cuarentena y el crecimiento epidémico es $R_0 < 4$, evidenciando de falta de identificabilidad en el modelo cuando se considera la dinámica de la cuarentena. Entonces, tomaría con cautela los periodos de cuarentena estudiados en los modelos. Ahora, ampliando el análisis a la identificabilidad práctica en las fases iniciales, encontramos una pérdida importante de identificabilidad de los parámetros, comparado con los resultados estructurales. Verificando tal vez la conclusión que se muestra en la referencia [139] que indica la no identificabilidad en los momentos previos al peak. Pero nos interesa analizar el período inicial para recuperar el número de reproducción con control R_0^c . Por lo que consideramos esencial aplicar nuestra metodología en diferentes períodos epidémicos para evaluar los parámetros de nuestro modelo. En nuestros experimentos, confirmamos que el parámetro β se recupera, y en los casos en que también se estiman las condiciones iniciales, la identificabilidad es una garantía para curvas con un crecimiento de $t > 20$ días. Esto, muestran otra característica interesante, la dependencia entre los tiempos de ajuste y las tasas de crecimiento, además de los períodos de cuarentena, que también afectan la identificabilidad, pero mantienen el parámetro β identificable. Finalmente, recogiendo las conclusiones del experimento con datos sintéticos, seleccionamos los conjuntos de parámetros (β), y $(\beta,\ J(0),\ I_s(0))$ para periodos sin cuarentena, y (β, p), y $(\beta, p, J(0), I_s(0))$ para periodos con cuarentena, para ajustarse a datos chilenos de algunas regiones. Encontramos que los parámetros del modelo no son identificables en las regiones chilenas con período de cuarentena en la fase inicial. Por otro lado, las regiones sin periodos de cuarentena tienen mejores resultados, donde entre más puntos de datos se ajustan mejores estimaciones. Para trabajos futuros, vemos la necesidad de seleccionar los tiempos adecuados para cada región chilena debido a que cada una tiene declarada una cuarentena particular; además, la calidad de los datos también es variable. Luego podemos comparar los tiempos de ajuste expuestos aquí [74] y comparar los resultados. Actualmente, los valores del parámetro β para la Región 8 en la fase inicial son cercanos.

Trabajo Futuro

En lineas generales podemos indicar que algunos escenarios de interés para futuras investigaciones son:

- Aumentar el estudio comparativo de PGMs a otros modelos no considerados y que pueden ser útiles en la epidemiología matemática, u otra línea de aplicación.
- Estructurar matemáticamente el concepto de distancia entre PGMs, que permitan caraterizar mejor tales modelos para el ajuste de enfermedades infecciosas.
- Por lo visto hasta ahora para el estudio del COVID-19 en Chile, se hace necesario considerar periodos de cuarentena en una linea de tiempo discreta, para recuperar de una forma más confiable en número de reproducción.
- Incluir un estudio de sensibilidad para los parámetros esitmados, que nos permita tener una conocimiento sobre las dependencias entres estos.
- Profundizar el estudio para establecer una metodología generalizada para el análisis computacional de la identificabilidad de parámetros en modelos compartimentales, donde se involucren nuevos procesos para mejorar los tiempos de ejecución del Bootstrapping, así como nuevos algoritmos para la estimación de parámetros, como lo puede ser las técnicas de Monte Carlo con cadenas de Markov o el reciente Monte Carlo hamiltoniano.

APPENDIX A

MATLAB Programs

Several numerical calculations contained in this work have been performed by using MATLAB routines, in particular, to implement Simulated Annealing we used different MATLAB functions, such as SIMULANNEALBND, LHSDESIGN, and ODE23S. In the following we expose the codes used to compute the parameter estimation, errors, and the plots presented in Section 2.3.2.

```
clear
close all
 for j=1:7
    %Calling to data curves generated with a growth model (model A) using 7 different ↩
        selection of parameter p
    load(sprintf('Curves-(% d).mat',j))
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    % Data curves are generated replacing the model A solution
    % with the parameters summarized in Table 2.
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

    %Data identification
    X0=data(1,2); timevect=data(:,1); CURVES=data(:,2);
    %Generating random 10 sets for the initial parameters(theta_1,theta_2,theta_3)
    lhs1=lhsdesign(10,3); %Latin hypercube sample of 10 values on each of 3 variables, ↩
        assuming in this case 0<theta_1,theta_2<3, and  0<theta_3<1

    for i=1:3
        theta_1s=3*lhs1(i,1); theta_2s=1*lhs1(i,2); theta_3s=lhs1(i,3);
        theta1=[theta_1s theta_2s theta_3s];

        %Fit data curves using another Growth model (Model B)
        LB=[0,0,0]; UB=[3,3,1];%Defining Lower and Upper Boundaries for each parameter of ↩
            model B
        %Implementation of Simulated Annealing algorithm to function objective to parameter ↩
            estimation of model B
        ObjectiveFunction = @(x) min_func(x,timevect,CURVES,X0,j);
        [P,FVAL,EXITFLAG,OUTPUT] = simulannealbnd(ObjectiveFunction,theta1,LB,UB);

        %Parameter Estimation results
        theta_hat=[];
```

```
         theta_hat(1)=P(1); theta_hat(2)=P(2); theta_hat(3)=P(3);

         %Compute the incidence curve to Model B with theta_hat
         [t, G]=slnModelB(theta_hat,timevect,X0);
         Incidence(:,i)=G;
         %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
         % As the GGoM, RM and LM models have explicit solutions,
         % the function slnModelB for these situations correspond
         % to their solution expression, but to GLM we used the
         % MATLAB function ODE23s to solve their ODE equation.
         %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

         %Compute errors
         RMSEB(:,i)=sqrt(mean((Incidence(:,i)-CURVES).^2));
         SSEB(:,i)=sum((Incidence(:,i)-CURVES).^2);
         MAEB(:,i)=mean(abs(Incidence(:,i)-CURVES));

         %Generated results
         Phats(i,:,j)=P;
         InitialParameter(i,:,j)=theta1;

         %Saved results
         save(sprintf('Incidences-(%d).mat',i), 'Incidence')
     end
     save(sprintf('RMSE-(%d).mat',j), 'RMSEB')
     save(sprintf('SSE-(%d).mat',j), 'SSEB')
     save(sprintf('MAE-(%d).mat',j), 'MAEB'
 end
 save('ParametersEstimation.mat', 'Phats')
 save('InitialParameters.mat', 'InitialParameter')

   %% Function objective
 function Z = min_func(p,t,CURVES,X0,i)
  [t,CP]=slnModelB(p,t,X0);
  Z = sum((CP-CURVES).^2,1).^(1/2);
 end

%% Compute of errors, parameters, and incidence for the smaller RMSEs
clear
load('ParametersEstimation.mat', 'Phats')
for k=1:7
    load(sprintf('RMSE-(%d).mat',k), 'RMSEB')
    load(sprintf('SSE-(%d).mat',k), 'SSEB')
    load(sprintf('MAE-(%d).mat',k), 'MAEB')
    load(sprintf('Incidences-(%d).mat',k), 'Incidence')
    %coordinates to smaller RMSE
    [~,r]=find(RMSEB==min(RMSEB));
    RMSEMin(:,k)=RMSEB(r);
    SSEMin(:,k)=SSEB(r);
    MAEMin(:,k)=MAEB(r);
    CoordinatesRMSE(:,:,k)=r;
    %Selection of Parameter estimation and incidence with smaller RMSE
    ParameterEstimationMin(k,:)=Phats(r,:,k);
    IncidencesMIN= Incidence(:,r);

    save(sprintf('IncidencesMIN-(%d).mat',k),'IncidencesMIN')
end
save('ParametersEstimationMIN.mat','ParameterEstimationMin')

%% Plotting  of curves and their fits showed in Figures 4, 6, 7 and 8
clear
close all
```

```
p=[1,0.995,0.99,0.98,0.95,0.85,0.8];
figure(1)
for i=1:7
    load(sprintf('Curves-(%d).mat',i))
    load(sprintf('IncidencesMINB-(%d).mat',i),'IncidencesMIN')
    hold on
    subplot(1,7,i)
    plot(data(:,1),data(:,2),'k*','LineWidth',2)
    hold on
    plot(data(:,1),IncidencesMIN,'r','LineWidth',2)
    set(gca,'FontSize', 10);
    xlabel('t','Fontsize',10);
    axis([0 inf 0 65]);
    set(gca,'fontsize',12);
    title(num2str(p(i)))
    xticks([0 5 10 15 20 30 40 50 60 ])
end
suptitle('Fit with Model B to Model A curves')
set(gcf,'color','white')
fig = gcf;
fig.Units = 'pixels';
fig.Position = [1 1 2000 400];

%% Calculation of the solution to GLM
%ODE model definition
function dx=GLM(t,x,r,p,k)
dx=r*(1-(x/k)).*x.^p;
end
%Application numerical method to approximate the solution of model GLM
function [t,CP]=slnGLM(P,t,X0)
%Parameters definition
r=P(1); p=P(2); K=P(3);
[r p K];
%Application of method to solve ODE
[t,x]=ode23s(@GLM,t,X0,[],r,p,K);
%Generation incidence curve
CP=[x(1);diff(x)];
end
```

APPENDIX B

Tables and Figures

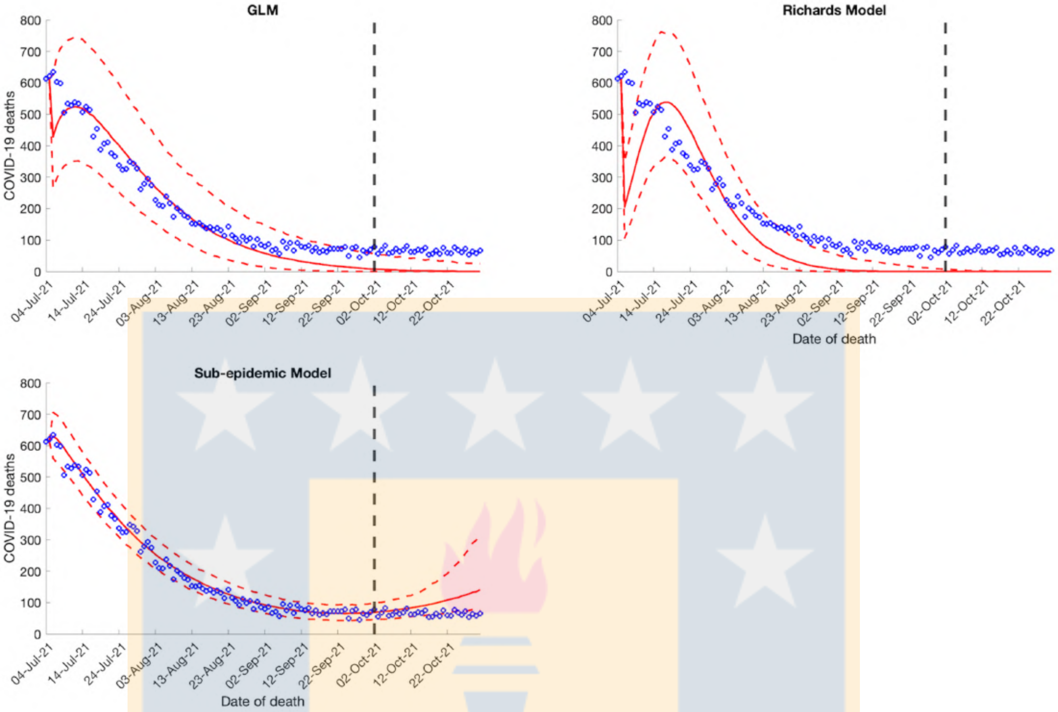


Figure B.1: 30-days ahead forecast of the national COVID-19 mortality curve in Colombia by calibrating the Richards, GLM and sub-epidemic wave model from July 04, 2021 to October 01, 2021.Blue circles correspond to the data points; the solid red line indicates the best model fit, and the red dashed lines represent the 95% PI. The vertical black dashed line represents the time of the start of the forecast period. The figure is taken from the published document. [150]

Model	RMSE	MAE	MIS	95% PI	WIS
GLM	14.37	29,8	225,95	93,33	23,03
Richards model	57,31	79,99	1365,8	42,22	62,44
Sub-epidemic wave model	18,08	13.4	80.06	98.88	8.52

Table B.1: Comparison of model performance metrics by calibrating the GLM, RM and the sub-epidemic model for 90 days of mortality data (July 4, 2021 to October 1, 2021). Higher 95%PI coverage and lower RMSE, MAE, WIS and MIS represent better performance. We highlight best performing model with green color.

Model	RMSE	MAE	MIS	95% PI	WIS
GLM	58,79	58,26	1529,1	0	54,21
Richards model	66,04	65,58	5340,1	0	66
Sub-epidemic wave model	46.59	38.3	296.8	60	23.9

Table B.2: Comparison of 30-day ahead forecasting performance (October 2, 2021 to October 31, 2021) by calibrating the GLM, RM and the sub-epidemic model for 90 days of mortality data (July 4, 2021 to October 1, 2021). Higher 95% PI coverage and lower RMSE, MAE, WIS and MIS represent better performance. We highlight best performing model with green color.
.

Par. Est.	R_0^c ref.	Experiments applied to each fit times																							
		1			2			3			4			5			6			7			8		
		20	30	40	20	30	40	20	30	40	20	30	40	20	30	40	20	30	40	20	30	40	20	30	40
α	3	▒	▒	▒	🟩	🟩	🟩	▒	▒	▒	🟩	🟩	🟩	▒	▒	▒	🟥	🟥	🟥	▒	▒	▒	🟥	🟥	🟥
	4	▒	▒	▒	🟩	🟩	🟩	▒	▒	▒	🟩	🟩	🟩	▒	▒	▒	🟥	🟥	🟥	▒	▒	▒	🟥	🟥	🟥
	5	▒	▒	▒	🟩	🟩	🟥	▒	▒	▒	🟩	🟩	🟥	▒	▒	▒	🟥	🟥	🟥	▒	▒	▒	🟥	🟥	🟥
β	3	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟥	🟥	🟩	🟥	🟩	🟩	🟥	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥
	4	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟥	🟩	🟩	🟥	🟩	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
	5	🟩	🟩	🟩	🟩	🟩	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥	🟩	🟩	🟥	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥
ρ_s	3	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥
	4	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥
	5	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥
$I_s(0)$	3	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
	4	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
	5	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
$J(0)$	3	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
	4	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
	5	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
R_0	3	🟩	🟩	🟩	🟥	🟩	🟩	🟥	🟩	🟩	🟥	🟩	🟩	🟥	🟩	🟩	🟥	🟩	🟩	🟥	🟥	🟩	🟥	🟩	🟩
	4	🟩	🟩	🟩	🟥	🟥	🟥	🟥	🟩	🟩	🟥	🟩	🟩	🟥	🟩	🟩	🟥	🟩	🟩	🟥	🟥	🟥	🟥	🟥	🟥
	5	🟩	🟩	🟩	🟥	🟥	🟥	🟥	🟩	🟩	🟥	🟥	🟥	🟥	🟩	🟩	🟥	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥

Table B.3: Summarize of results obtained for Scenario 1. The green box represents that the parameter estimated satisfies the parameter identifiability conditions, i.e., this has small interval confidence with mean value-centered and a low MSE. While the red box, the parameter does not satisfy the identifiability criteria

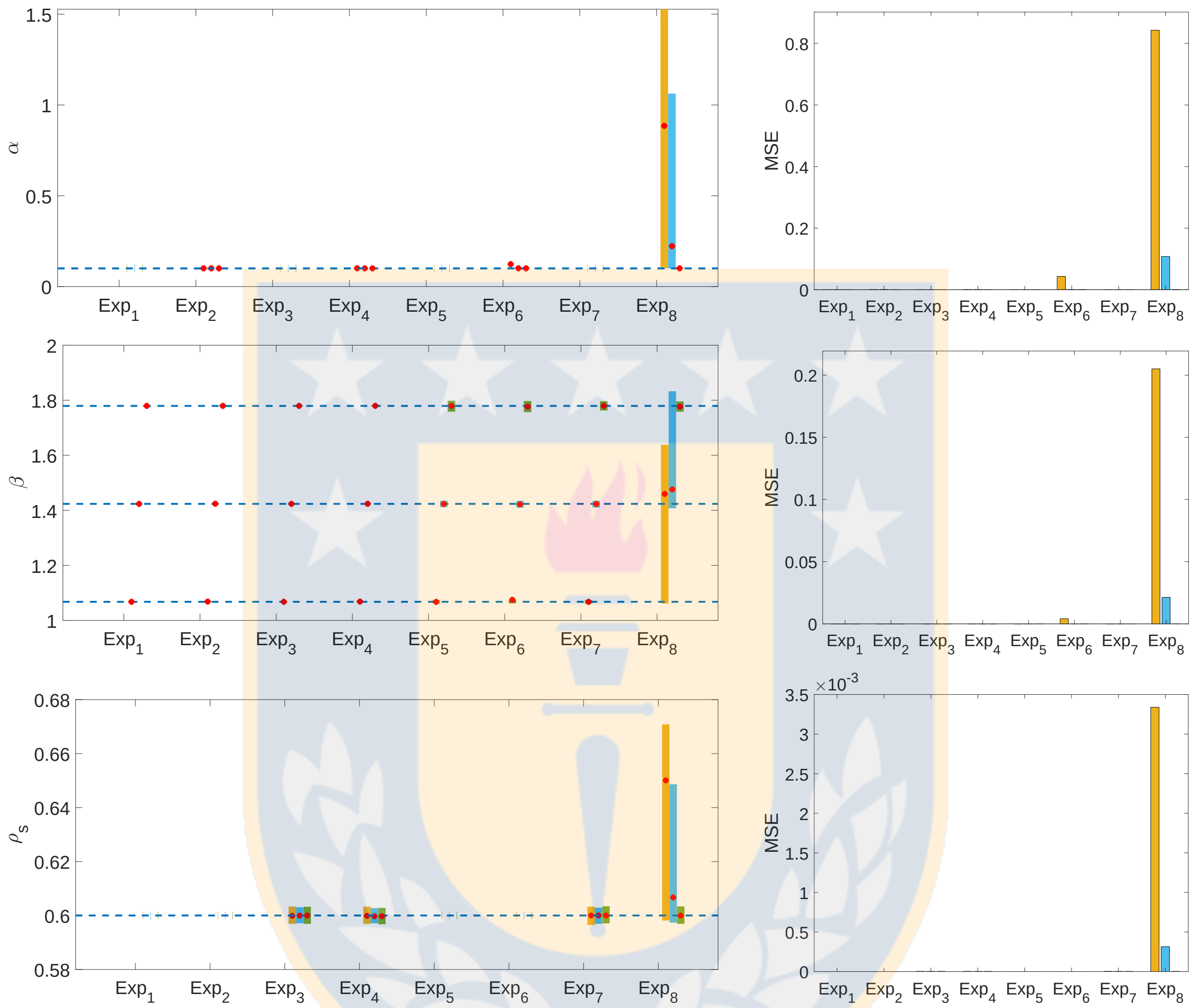


Figure B.2: Summary of results using conditions of experiment of scenario 1 (Parameters (α, β, ρ_s)). Here and in Figures B.4, B.5 and B.6 in per row, the left plot corresponds to the 95% confidence intervals (vertical lines) for the distributions of each estimated parameter obtained through 250 realizations of the synthetic data generated for Scenario 1. Each red point denotes the mean estimated parameter value. The light-blue dashed horizontal line represents the true value (or assumed) for each parameter estimated; the experiments applied to each parameter are fixed on the x-axis. The right plot corresponds to the mean squared error (MSE) distribution of parameter estimates considering each experiment and dataset. Finally, Yellow, blue, and green lines or bars are grouped for each experiment and represent each R_0^c assumed for creating the synthetic data, i.e., $R_0^c = 3$, 4, and 5, respectively.

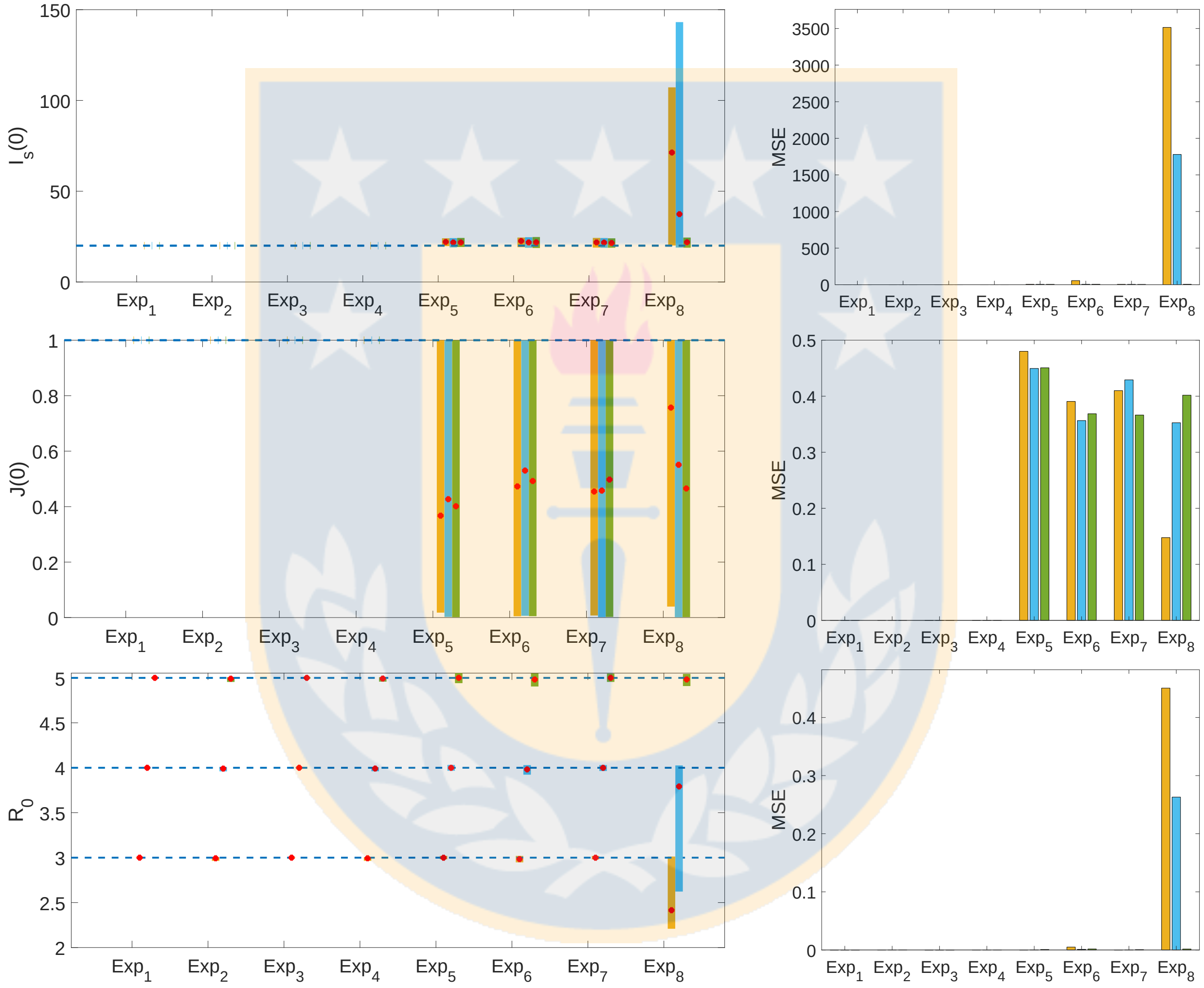


Figure B.3: Summary of results using conditions of scenario 1 (Parameters $(I_s(0),\, J(0),\, R_0^c)$).

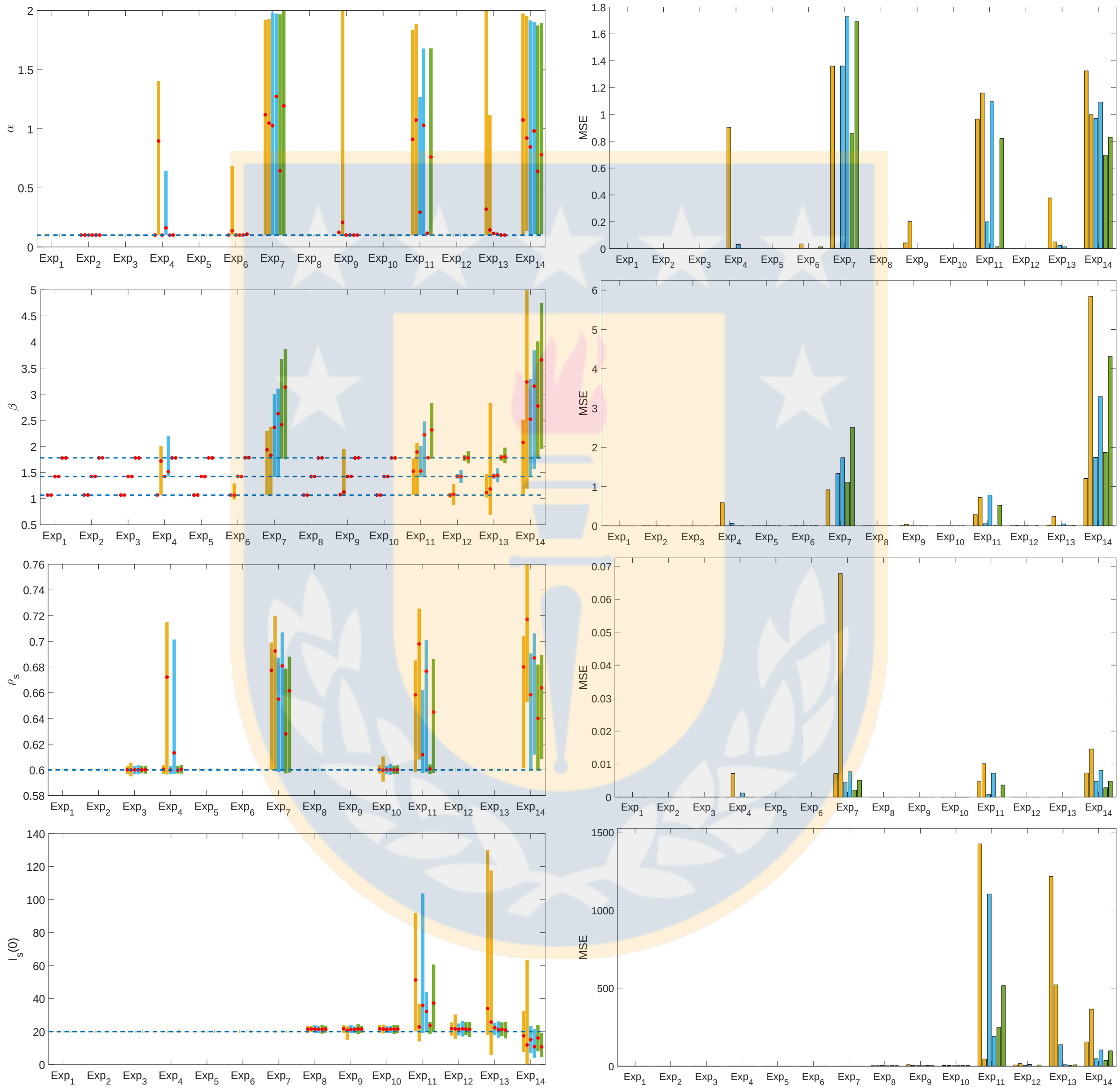


Figure B.4: Summary of results using conditions of scenario 2 (Parameters (α, β, ρ_s, $I_s(0)$)).

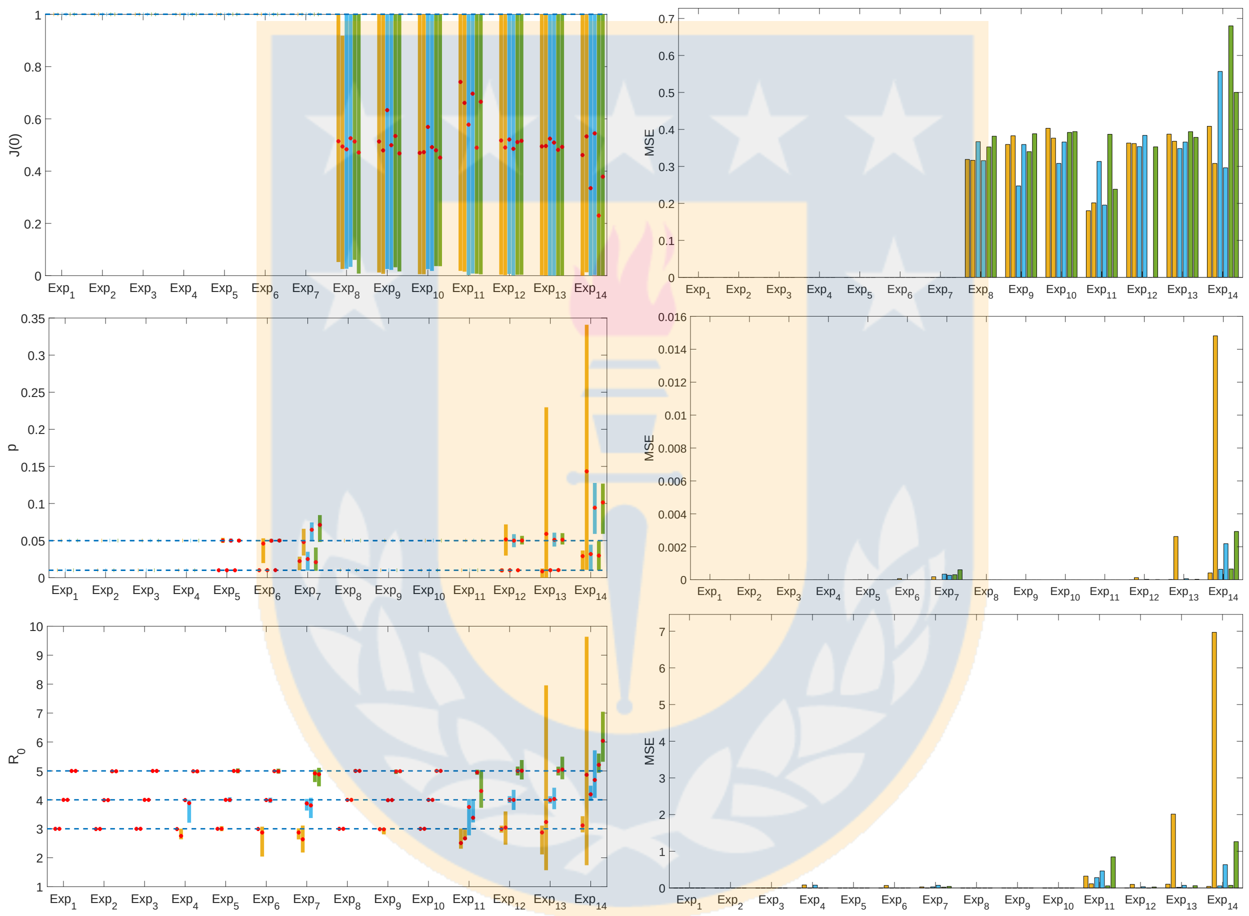
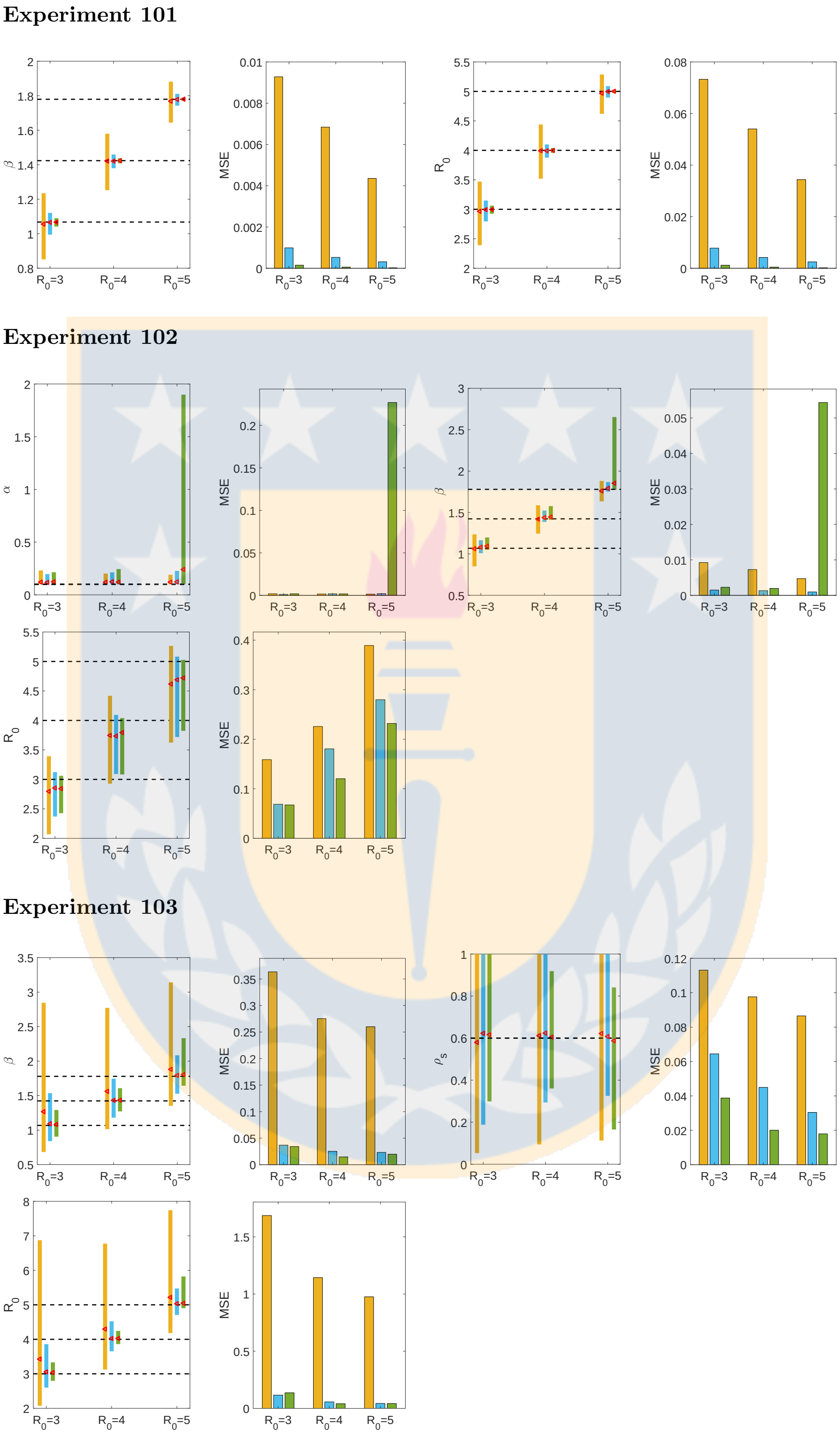


Figure B.5: Summary of results using conditions of scenario 2 (Parameters ($J(0)$, p, R_0^c)).



Experiment 101

Experiment 102

Experiment 103

Figure B.6: Summary of results to practical identifiability using conditions of scenario 1 (Experiments 101- 102 - 103).

Par.	p	R_0	Experiments applied to each fit times																																							
Est.	ref.	ref.	1			2			3			4			5			6			7			8			9			10			11			12			13			14		
			20	30	40	20	30	40	20	30	40	20	30	40	20	30	40	20	30	40	20	30	40	20	30	40	20	30	40	20	30	40	20	30	40	20	30	40	20	30	40	20	30	40
α	0.01	3	▒	▒	▒	🟥	🟩	🟩	▒	▒	▒	🟥	🟩	🟩	▒	▒	▒	🟩	🟩	🟩	🟥	🟥	🟥	▒	▒	▒	🟥	🟥	🟥	▒	▒	▒	🟥	🟥	🟥	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥
		4	▒	▒	▒	🟩	🟥	🟥	▒	▒	▒	🟥	🟥	🟥	▒	▒	▒	🟩	🟩	🟩	🟥	🟥	🟥	▒	▒	▒	🟥	🟥	🟥	▒	▒	▒	🟥	🟥	🟥	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥
		5	▒	▒	▒	🟩	🟥	🟩	▒	▒	▒	🟥	🟩	🟥	▒	▒	▒	🟩	🟩	🟥	🟥	🟥	🟥	▒	▒	▒	🟥	🟥	🟥	▒	▒	▒	🟥	🟥	🟥	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥
	0.05	3	▒	▒	▒	🟥	🟩	🟥	▒	▒	▒	🟥	🟥	🟥	▒	▒	▒	🟥	🟩	🟥	🟥	🟥	🟥	▒	▒	▒	🟥	🟥	🟥	▒	▒	▒	🟥	🟥	🟥	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥
		4	▒	▒	▒	🟩	🟥	🟥	▒	▒	▒	🟥	🟥	🟥	▒	▒	▒	🟩	🟥	🟥	🟥	🟥	🟥	▒	▒	▒	🟥	🟥	🟥	▒	▒	▒	🟥	🟥	🟥	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥
		5	▒	▒	▒	🟩	🟥	🟥	▒	▒	▒	🟥	🟥	🟥	▒	▒	▒	🟩	🟥	🟥	🟥	🟥	🟥	▒	▒	▒	🟥	🟥	🟥	▒	▒	▒	🟥	🟥	🟥	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥
β	0.01	3	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟥	🟩	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟩	🟩	🟥	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		4	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟥	🟩	🟥	🟥	🟩	🟥	🟥	🟩	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟩	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		5	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟩	🟩	🟥	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
	0.05	3	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟩	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		4	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟩	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		5	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟩	🟩	🟥	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
ρ_s	0.01	3	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥
		4	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥
		5	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥
	0.05	3	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥
		4	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥
		5	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥
I_s	0.01	3	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		4	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		5	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
	0.05	3	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		4	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		5	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
$J(0)$	0.01	3	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		4	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		5	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
	0.05	3	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		4	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		5	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
p	0.01	3	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		4	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟩	🟥	🟥	🟩	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		5	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟩	🟥	🟥	🟩	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
	0.05	3	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		4	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		5	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
R_0	0.01	3	🟩	🟩	🟩	🟥	🟩	🟩	🟥	🟩	🟩	🟥	🟩	🟩	🟥	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟩	🟩	🟩	🟩	🟩	🟥	🟥	🟩	🟥	🟩	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		4	🟩	🟩	🟩	🟥	🟩	🟩	🟥	🟩	🟩	🟥	🟩	🟩	🟥	🟥	🟩	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟩	🟩	🟥	🟩	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		5	🟩	🟩	🟩	🟥	🟥	🟥	🟥	🟩	🟩	🟥	🟩	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟩	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
	0.05	3	🟩	🟩	🟩	🟥	🟩	🟩	🟥	🟥	🟩	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟩	🟩	🟥	🟩	🟩	🟥	🟥	🟩	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		4	🟩	🟩	🟩	🟥	🟩	🟩	🟥	🟩	🟩	🟥	🟩	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟩	🟩	🟥	🟩	🟩	🟥	🟥	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
		5	🟩	🟩	🟩	🟥	🟥	🟩	🟥	🟩	🟩	🟥	🟩	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟩	🟩	🟥	🟥	🟩	🟥	🟩	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥

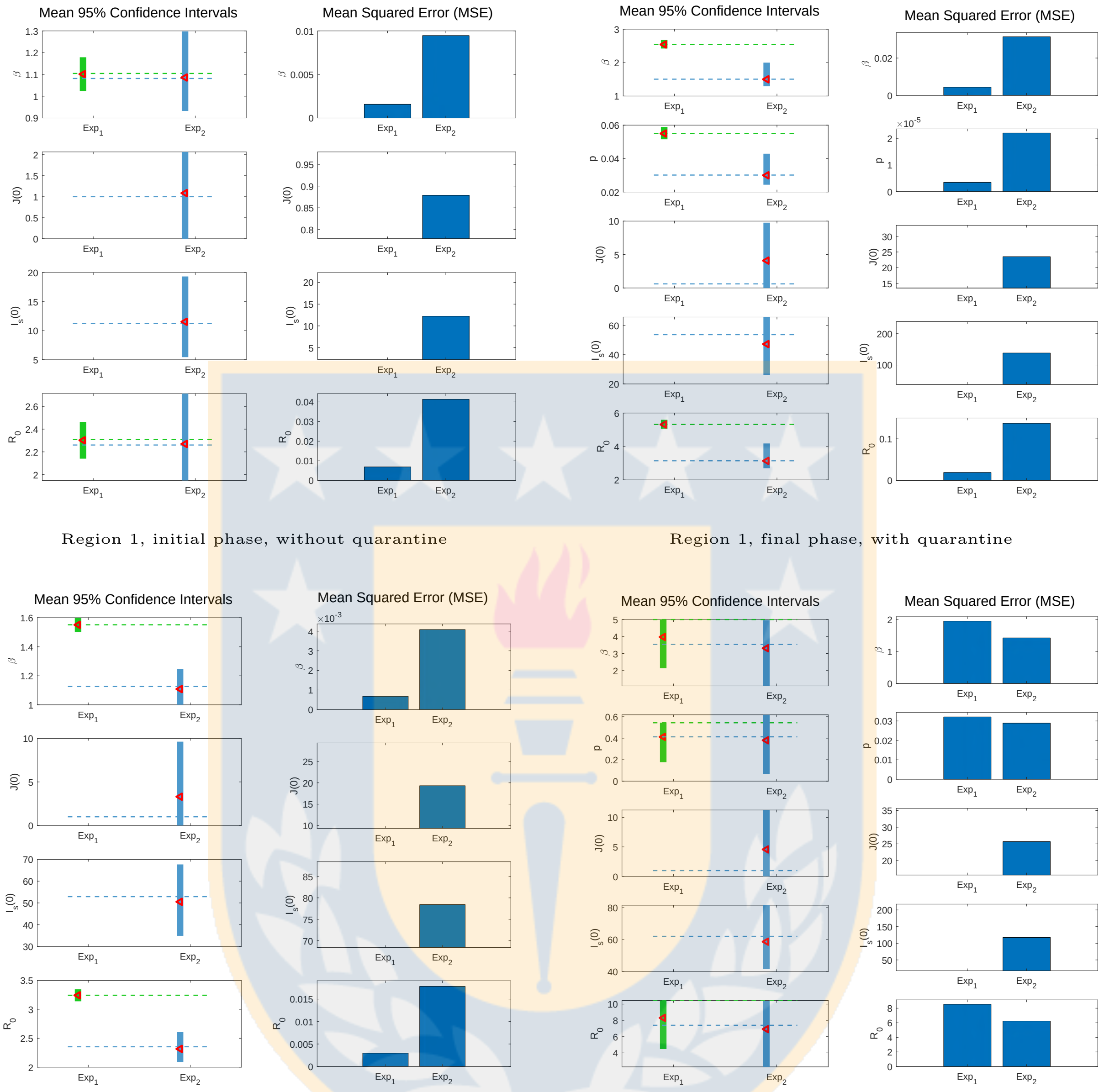
Table B.4: Summarize of results obtained for Scenario 2. The green box represents that the parameter estimated satisfies the parameter identifiability conditions, i.e., this has small interval confidence with mean value-centered and a low MSE. While the red box, the parameter does not satisfy the identifiability criteria.

Parameter	Selected for simulations	Source
h	0	Assumed
q_{e}	0.1	[37]
q_{a}	0.4	[37]
q	0.4	Assumed
$1/\kappa_1$	2.5	[120, 194]
$1/\kappa_2$	2.5	[92, 189]
ρ_{a}	0.4	[107, 115]
ρ_{s}	0.5	Assumed
$1/\alpha$	5	Assumed
$1/\gamma_1$	7	[120, 193]
$1/\gamma_2$	5	[154]
δ	0.021	National death case data

Table B.5: Fixed parameters for running experiments.

Región	Initial phase	Final phase
	$1/\lambda$	$1/\lambda$
1	0	67
8	0	11
10	21	-

Table B.6: Duration of quatantine periods.



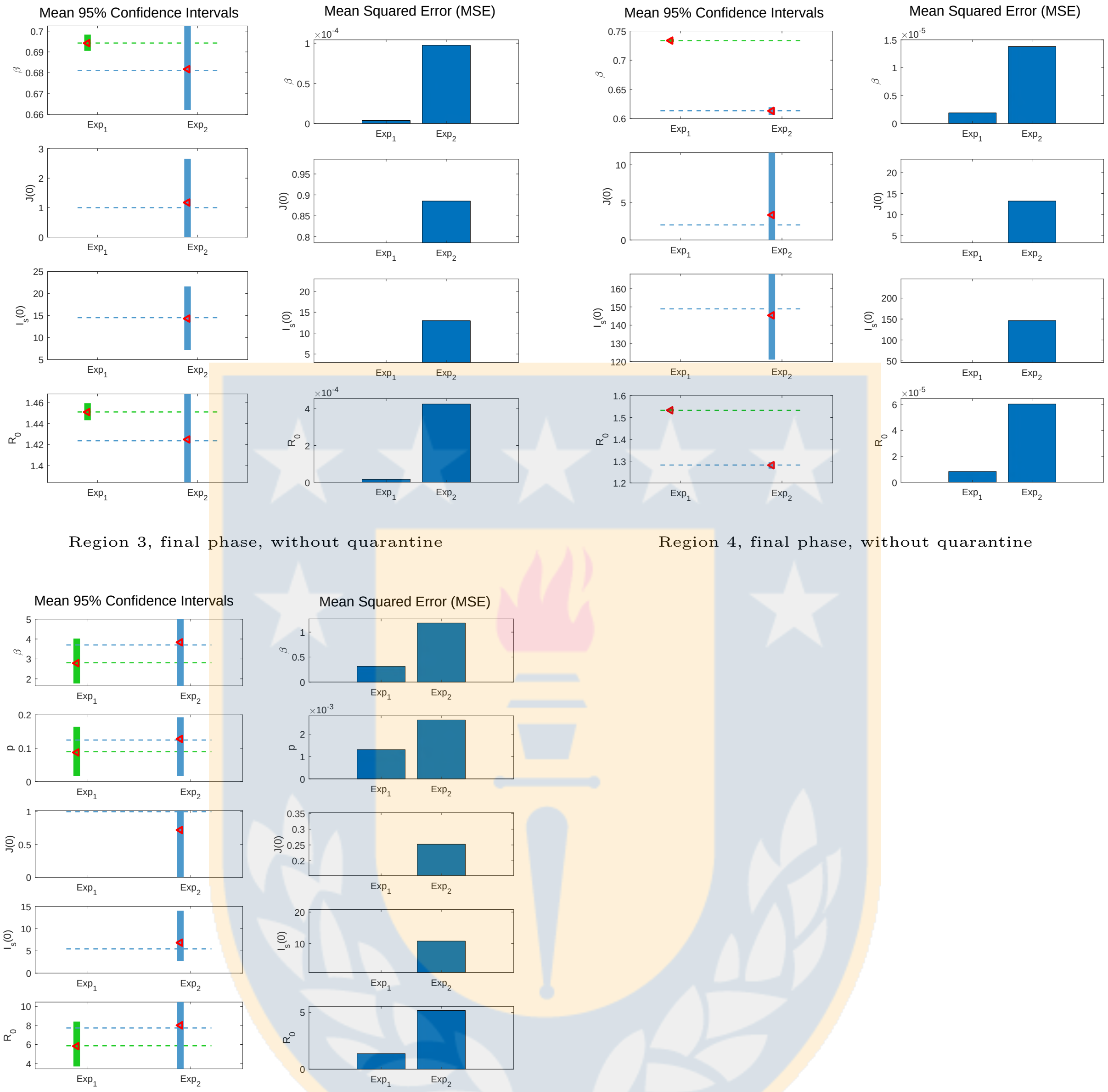
Region 1, initial phase, without quarantine

Region 1, final phase, with quarantine

Region 8, initial phase, without quarantine

Region 8, final phase, with quarantine

Figure B.7: Summary of results using regional Chilean data (Region 1 and 8). Here and in Figure B.9 the left plot corresponds to the 95% confidence intervals (vertical lines) for the distributions of each estimated parameter obtained through 250 realizations of the synthetic data generated for the best fit using the Chilean data. Each red left-pointing trianglet denotes the mean estimated parameter value. The dashed horizontal line represents the value estimated for each parameter in the best-fit. The experiments applied to each parameter set are fixed on the x-axis. The right plot corresponds to the mean squared error (MSE) distribution of parameter estimates considering each experiment. Finally, blue, and green lines or bars represent each experiment.



Region 3, final phase, without quarantine

Region 4, final phase, without quarantine

Region 10, initial phase, with quarantine

Figure B.8: Summary of results using regional Chilean data (Region 3, 4 and 10).

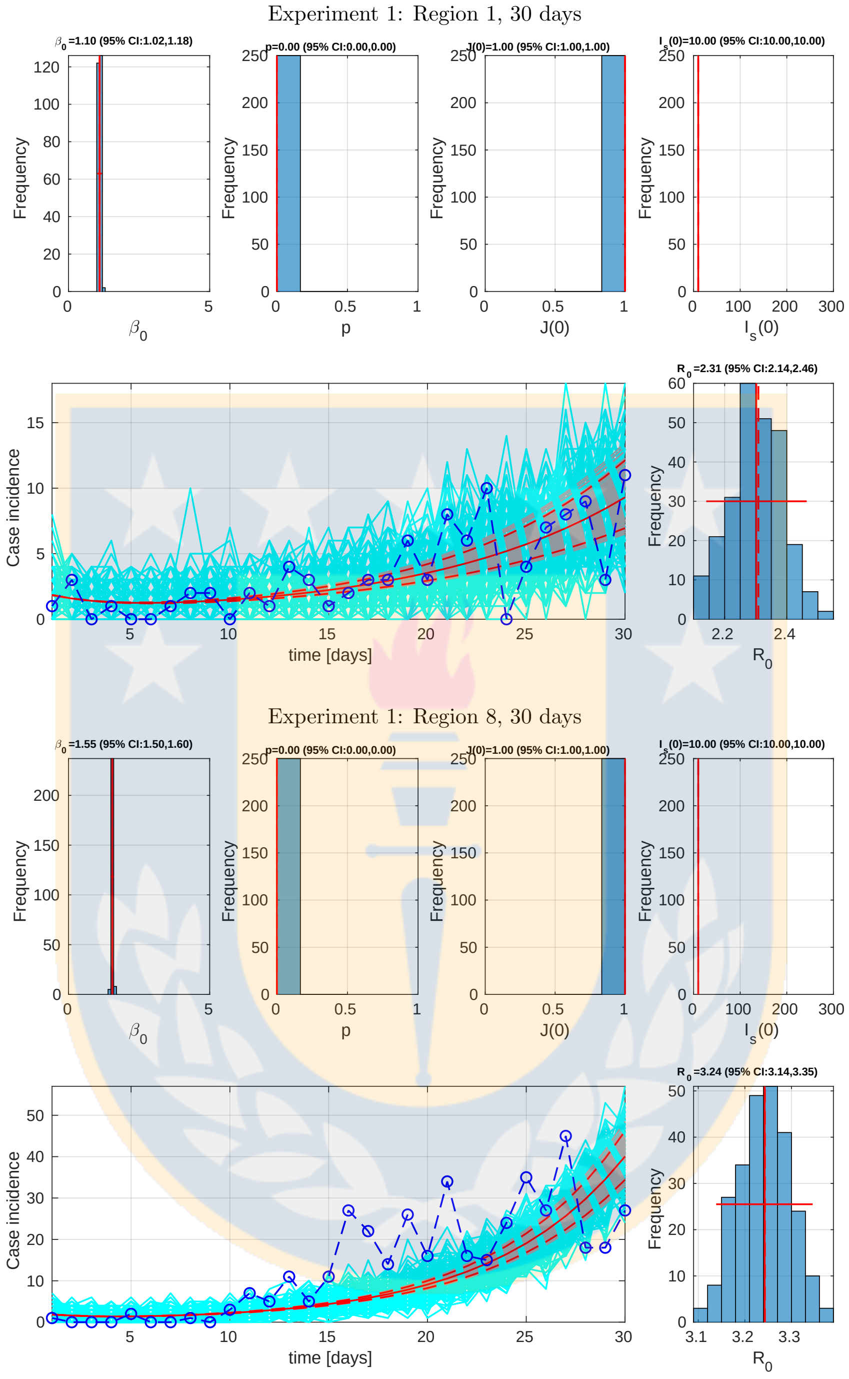


Figure B.9: Results of the bootstrapping process to the parameters estimated for Experiment 1 (β) for Regions 1 and 8, in the initial phase of 30 days. Here and Figures B.11-B.16, the histograms display the empirical distribution of the parameters estimated using 250 bootstrap realizations. Their confidence intervals appear as a horizontal red line, the mean value as a vertical red line, and the value estimated corresponds to a vertical dashed red line. The bottom panel with curves shows the fit of the COVID-19 model to the 200 days Chilean data. The blue line with circles is the daily data, while the solid red line corresponds to the best fit using the SA-LSQ between our model and the dataset. The dashed red lines correspond to the 95% confidence bands around the best fit of the model to the data. Observation: Some experiments have parameters fixed, then its histograms appear with values set.

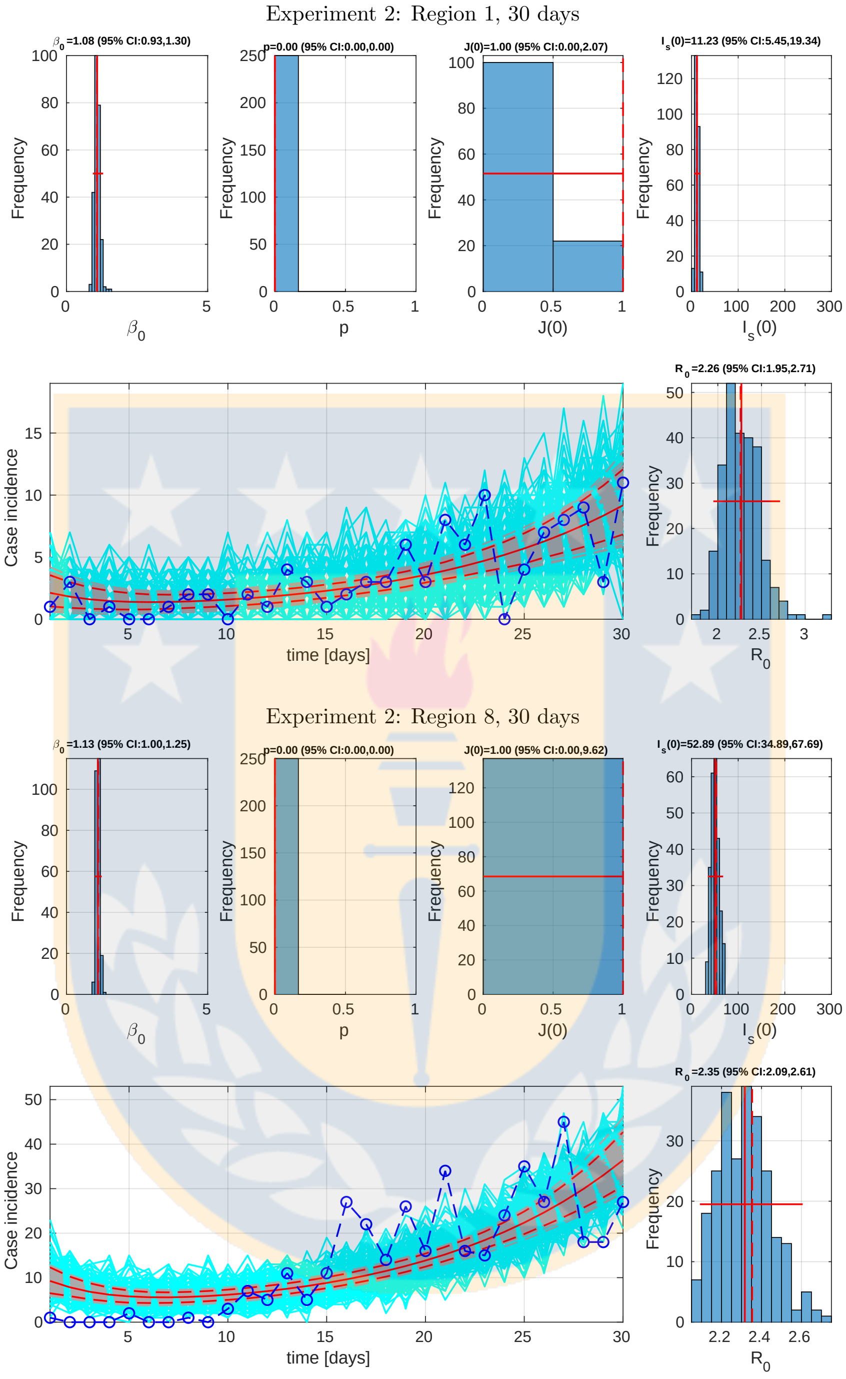
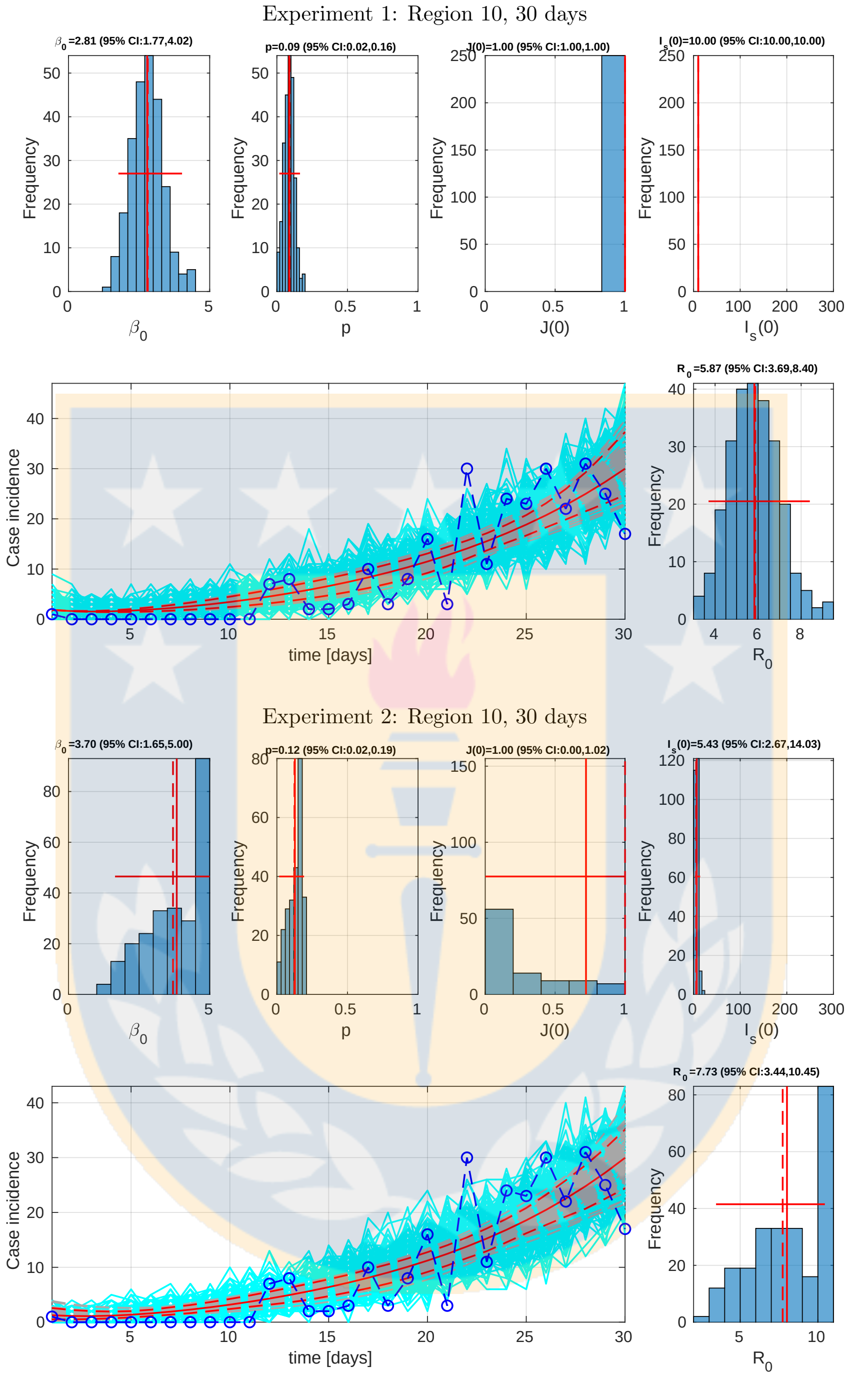


Figure B.10: Results of the bootstrapping process to the parameters estimated for Experiment 2 (β, $J(0)$, $I_s(0)$) for Regions 1 and 8, in the initial phase of 30 days.


Experiment 1: Region 10, 30 days

Experiment 2: Region 10, 30 days

Figure B.11: Results of the bootstrapping process to the parameters estimated for Experiments 1 and 2 with quarantine, applied to Region 10, in the initial phase of 30 days.

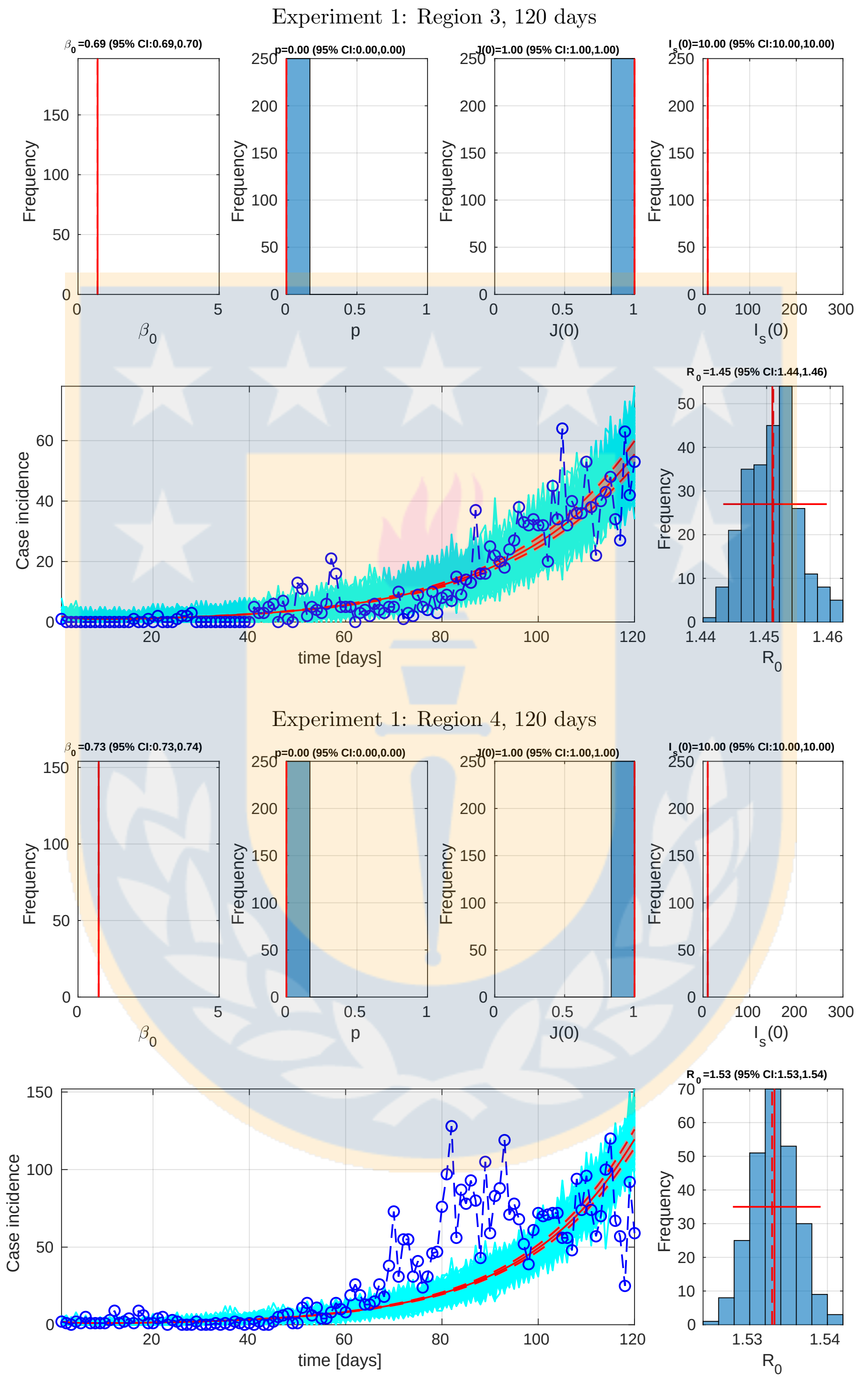
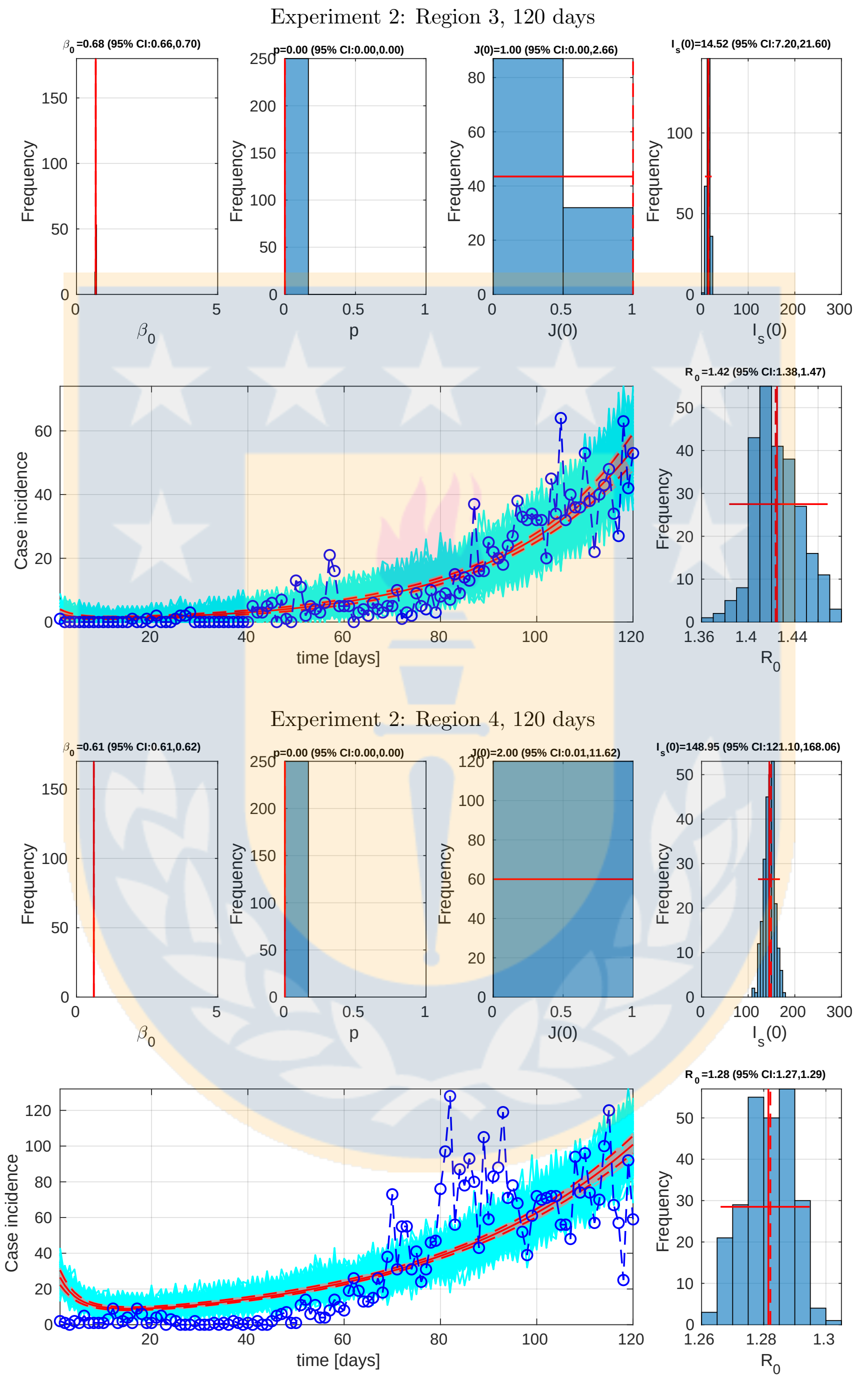


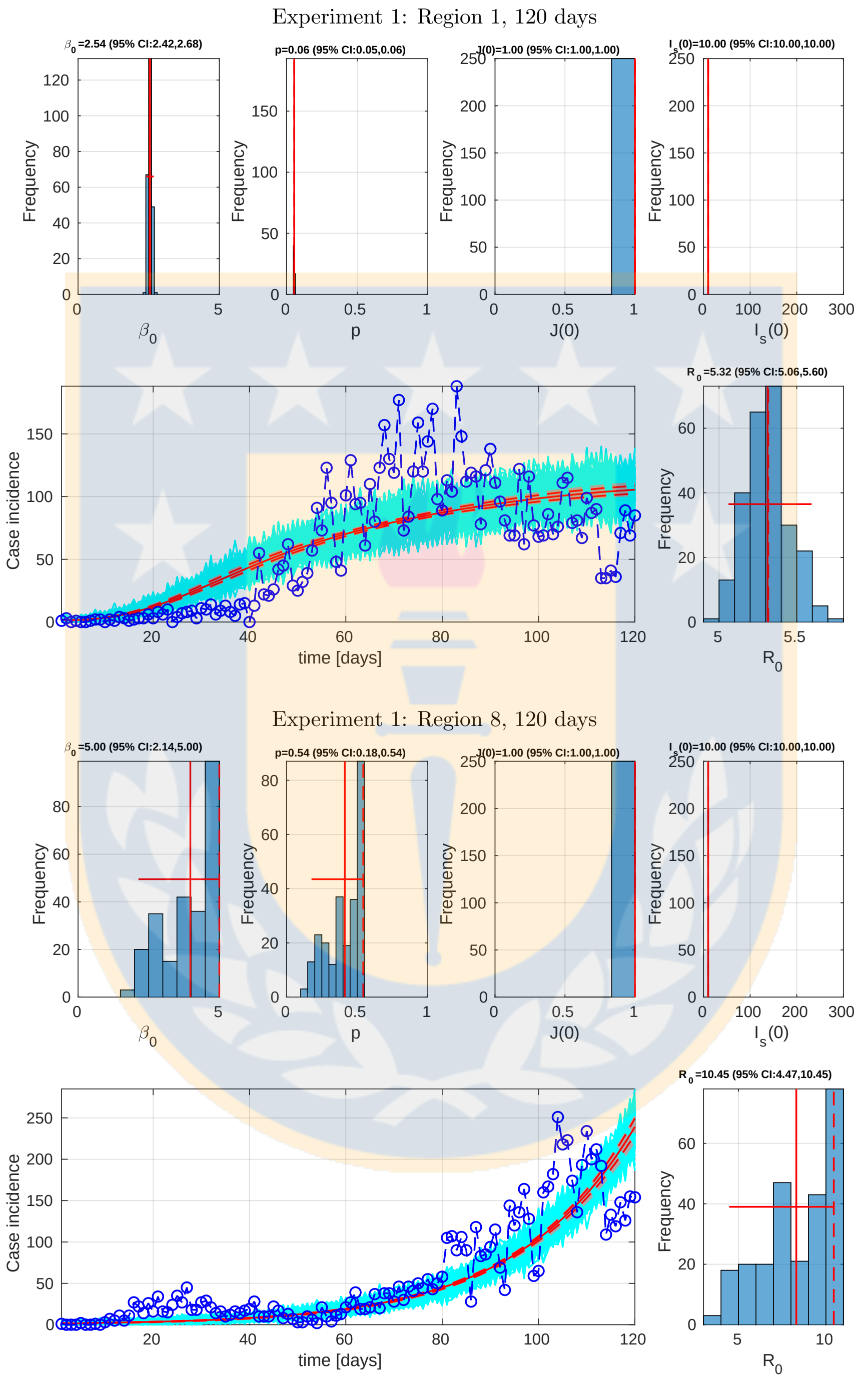
Figure B.12: Results of the bootstrapping process to the parameters estimated for Experiment 1 (β) for Regions 3 and 4, in the final phase of 120 days.



Experiment 2: Region 3, 120 days

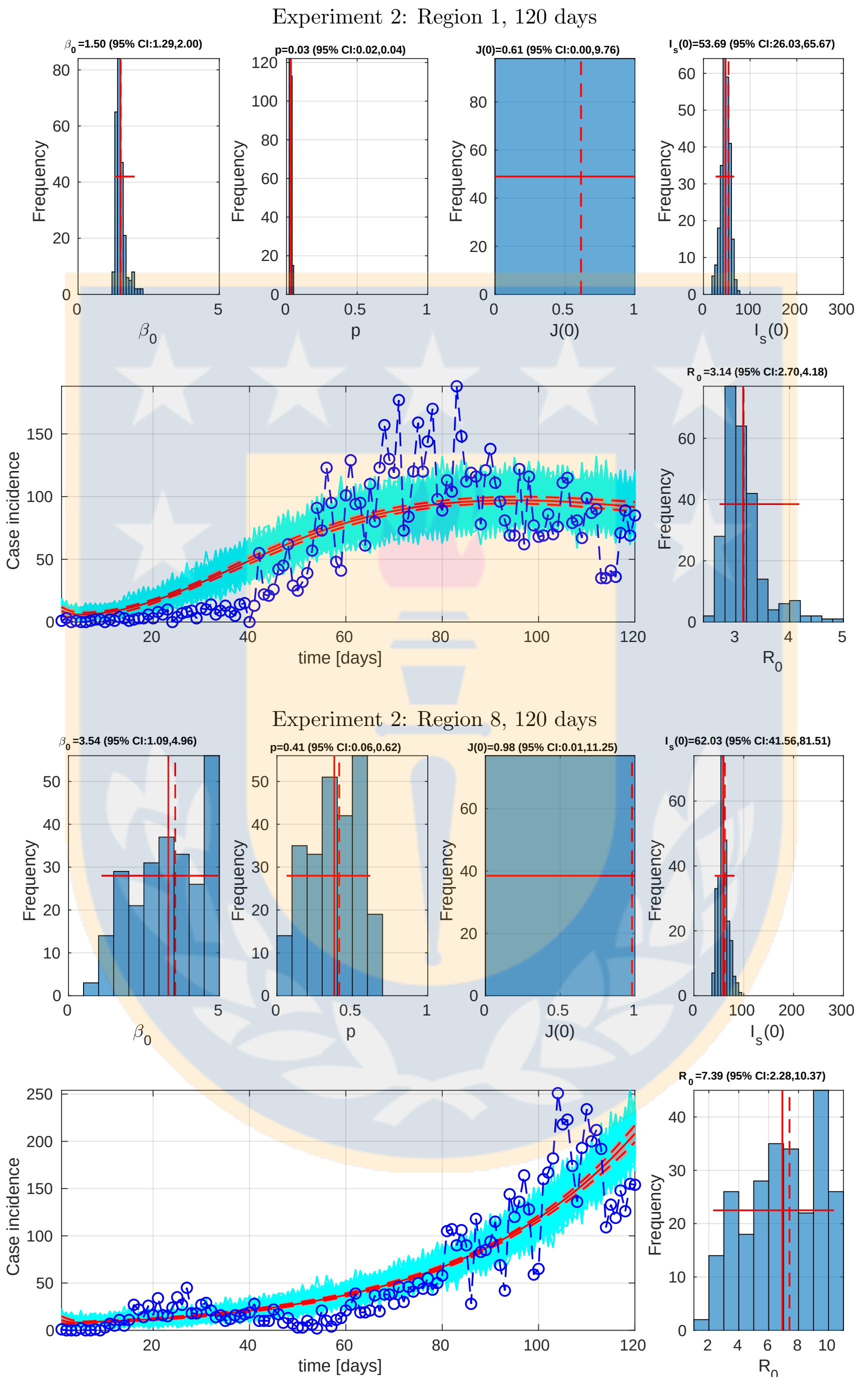
Experiment 2: Region 4, 120 days

Figure B.13: Results of the bootstrapping process to the parameters estimated for Experiment 2 (β, $J(0)$, $I_s(0)$) for Regions 3 and 4, in the final phase of 120 days.


Experiment 1: Region 1, 120 days

Experiment 1: Region 8, 120 days

Figure B.14: Results of the bootstrapping process to the parameters estimated for Experiment 1 (β, p) for Regions 1 and 8, in the final phase of 120 days.



Experiment 2: Region 1, 120 days

Experiment 2: Region 8, 120 days

Figure B.15: Results of the bootstrapping process to the parameters estimated for Experiment 2 (β, p, $J(0)$, $I_s(0)$) for Regions 1 and 8, in the final phase of 120 days.

References

[1] B. Abbasi, A. H. E. Jahromi, J. Arkat, and M. Hosseinkouchack, *Estimating the parameters of weibull distribution using simulated annealing algorithm*, Applied Mathematics and Computation, 183 (2006), pp. 85–93.

[2] R. M. Anderson and R. M. May, *Infectious Diseases of Humans, Dynamics and Control*, Oxford University Press, 1991.

[3] Anonymous, *XXII. the epidemiological observations made by the commission in bombay city*, Epidemiology and Infection, 7 (1907), pp. 724–798.

[4] J. Arellana, L. Márquez, and V. Cantillo, *Covid-19 outbreak in colombia: An analysis of its impacts on transport systems*, Journal of Advanced Transportation, (2020).

[5] A. R. Arenas, N. B. Thackar, and E. C. Haskell, *The logistic growth model as an approximating model for viral load measurements of influenza a virus*, Mathematics and Computers in Simulation, 133 (2017), pp. 206–222.

[6] J. Aroesty, T. Lincoln, N. Shapiro, and G. Boccia, *Tumor growth and chemotherapy: Mathematical methods, computer simulations, and experimental foundations*, Mathematical Biosciences, 17 (1973), pp. 243–300.

[7] H. T. Banks, S. Hu, and W. C. Thompson, *Modeling and inverse problems in the presence of uncertainty*, CRC Press, 2014.

[8] A. Barrera, P. Román-Román, and F. Torres-Ruiz, *A hyperbolastic type-i diffusion process: Parameter estimation by means of the firefly algorithm*, Biosystems, 163 (2018), pp. 11–22.

[9] M. A. Benítez, C. Velasco, A. R. Sequeira, J. Henríquez, F. M. Menezes, and F. Paolucci, *Responses to covid-19 in five latin american countries*, Health Policy and Technology, 9 (2020), pp. 525–559.

[10] C. BIRCH, *A new generalized logistic sigmoid growth equation compared with the richards growth equation*, Annals of Botany, 83 (1999), pp. 713–723.

[11] B. Bolker, *Measles times-series data. professor b. bolker's personal data repository (accessed 27 september 2016).*, tech. rep., McMaster University, 2016.

[12] J. Bracher, E. L. Ray, T. Gneiting, and N. G. Reich, *Evaluating epidemic forecasts in an interval format*, PLOS Computational Biology, 17 (2021), pp. 1–15.

[13] F. Brauer and C. Castillo-Chávez, *Mathematical Models in Population Biology and Epidemiology*, Springer New York, 2001.

[14] F. Brauer and C. Kribs, *Dynamical Systems for Biological Modeling*, Chapman and Hall/CRC, Dec. 2015.

[15] M. Braun, *Differential Equations and Their Applications*, Springer New York, 1993.

[16] R. Bürger, G. Chowell, and L. Y. Lara-Díaz, *Measuring differences between phenomenological growth models applied to epidemiology*, Mathematical Biosciences, 334 (2021), p. 108558.

[17] N. F. Britton, *Essential Mathematical Biology*, Springer London, 2003.

[18] S. P. Brooks and B. J. T. Morgan, *Optimization using simulated annealing*, The Statistician, 44 (1995), p. 241.

[19] M. Brown, C. Cain, J. Whitfield, E. Ding, S. Y. D. Valle, and C. A. Manore, *Modeling zika virus spread in colombia using google search queries and logistic power models*, bioRxiv, (2018).

[20] J. H. Buckner, G. Chowell, and M. R. Springborn, *Dynamic prioritization of covid-19 vaccines when social distancing is limited for essential workers*, Proceedings of the National Academy of Sciences, 118 (2021).

[21] B. Buonomo, P. Manfredi, and A. d'Onofrio, *Optimal time-profiles of public health intervention to shape voluntary vaccination for childhood diseases*, Journal of Mathematical Biology, 78 (2018), pp. 1089–1113.

[22] R. Bürger, G. Chowell, and L. Y. Lara-Díaz, *Comparative analysis of phenomenological growth models applied to epidemic outbreaks*, Mathematical Biosciences and Engineering, 16 (2019), pp. 4250–4273.

[23] T. Burki, *COVID-19 in latin america*, The Lancet Infectious Diseases, 20 (2020), pp. 547–548.

[24] T. Burki, *Understanding variants of sars-cov-2*, The Lancet, (2021).

[25] A. Camacho, A. Kucharski, S. Funk, J. Breman, P. Piot, and W. Edmunds, *Potential for large outbreaks of ebola virus disease*, Epidemics, 9 (2014), pp. 70–78.

[26] C. Castilho, J. A. M. Gondim, M. Marchesin, and M. Sabeti, *Assessing the efficiency of different control strategies for the coronavirus (covid-19) epidemic*, Mathematical and Statistical Estimation Approaches in Epidemiology, 2020 (2020), pp. 1–17.

[27] M. Catala, S. Alonso, E. A. Lacalle, D. Lopez, P.-J. Cardona, and C. Prats, *Empiric model for short-time prediction of COVID-19 spreading*, medRxiv, (2020).

[28] J. F.-W. Chan et al., *A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster*, The Lancet, 395 (2020), pp. 514–523.

[29] J. D. Chapman and N. D. Evans, *The structural identifiability of susceptible–infective–recovered type epidemic models with incomplete immunity and birth targeted vaccination*, Biomedical Signal Processing and Control, 4 (2009), pp. 278–284.

[30] O.-T. Chis, J. R. Banga, and E. Balsa-Canto, *Structural identifiability of systems biology models: A critical comparison of methods*, PLOS ONE, 6 (2011), pp. 1–16.

[31] D. Chowell, K. Roosa, R. Dhillon, G. Chowell, and D. Srikrishna, *Sustainable social distancing through facemask use and testing during the covid-19 pandemic*, medRxiv, (2020).

[32] G. Chowell, *Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts*, Infectious Disease Modelling, 2 (2017), pp. 379–398.

[33] G. Chowell, F. Abdirizak, S. Lee, J. Lee, E. Jung, H. Nishiura, and C. Viboud, *Transmission characteristics of mers and sars in the healthcare setting: a comparative study*, BMC Med, 13 (2015).

[34] G. Chowell, C. Ammon, N. Hengartner, and J. Hyman, *Transmission dynamics of the great influenza pandemic of 1918 in geneva, switzerland: Assessing the effects of hypothetical interventions*, Journal of Theoretical Biology, 241 (2006), pp. 193–204.

[35] G. Chowell and F. Brauer, *The basic reproduction number of infectious diseases: Computation and estimation using compartmental epidemic models*, in Mathematical and Statistical Estimation Approaches in Epidemiology, Springer Netherlands, 2009, pp. 1–30.

[36] G. Chowell et al., *Using phenomenological models to characterize transmissibility and forecast patterns and final burden of zika epidemics*, PLoS Currents, (2016).

[37] G. Chowell, P. Fenimore, M. Castillo-Garsow, and C. Castillo-Chavez, *SARS outbreaks in ontario, hong kong and singapore: the role of diagnosis and isolation as a control mechanism*, Journal of Theoretical Biology, 224 (2003), pp. 1–8.

[38] G. Chowell, N. Hengartner, C. Castillo-Chavez, P. Fenimore, and J. Hyman, *The basic reproductive number of ebola and the effects of public health measures: the cases of congo and uganda*, Journal of Theoretical Biology, 229 (2004), pp. 119–126.

[39] G. Chowell, H. Nishiura, and L. M. Bettencourt, *Comparative estimation of the reproduction number for pandemic influenza from daily case notification data*, Journal of The Royal Society Interface, 4 (2006), pp. 155–166.

[40] G. Chowell, A. Rivas, N. Hengartner, J. Hyman, and C. Castillo-Chavez, *The role of spatial mixing in the spread of foot-and-mouth disease*, Preventive Veterinary Medicine, 73 (2006), pp. 297–314.

[41] G. Chowell, A. L. Rivas, S. D. Smith, and J. M. Hyman, *Identification of case clusters and counties with high infective connectivity in the 2001 epidemic of foot-and-mouth disease in uruguay*, American Journal of Veterinary Research, 67 (2006), pp. 102–113.

[42] G. Chowell, L. Sattenspiel, S. Bansal, and C. Viboud, *Mathematical models to characterize early epidemic growth: A review*, Physics of Life Reviews, 18 (2016), pp. 66–97.

[43] G. Chowell, E. Shim, F. Brauer, P. Diaz-Dueñas, J. M. Hyman, and C. Castillo-Chavez, *Modelling the transmission dynamics of acute haemorrhagic conjunctivitis: application to the 2003 outbreak in mexico*, Statistics in Medicine, 25 (2006), pp. 1840–1857.

[44] G. Chowell, A. Tariq, and J. M. Hyman, *A novel sub-epidemic modeling framework for short-term forecasting epidemic waves*, BMC Medicine, 17 (2019).

[45] G. Chowell and C. Viboud, *Is it growing exponentially fast? impact of assuming exponential growth for characterizing and forecasting epidemics with initial near-exponential growth dynamics*, Infectious Disease Modelling, 1 (2016), pp. 71–78.

[46] G. Chowell, C. Viboud, J. M. Hyman, and L. Simonsen, *The western africa ebola virus disease epidemic exhibits both global exponential and local polynomial growth rates*, PLoS Currents, (2014).

[47] G. Chowell, C. Viboud, L. Simonsen, S. Merler, and A. Vespignani, *Perspectives on model forecasts of the 2014–2015 ebola epidemic in west africa: lessons and the way forward*, BMC Medicine, 15 (2017).

[48] G. Chowell, C. Viboud, L. Simonsen, and S. M. Moghadas, *Characterizing the reproduction number of epidemics with early subexponential growth dynamics*, Journal of The Royal Society Interface, 13 (2016), p. 20160659.

[49] G. Chowell, C. Viboud, X. Wang, S. M. Bertozzi, and M. A. Miller, *Adaptive vaccination strategies to mitigate pandemic influenza: Mexico as a case study*, PLOS ONE, 4 (2009), pp. 1–9.

[50] I. Chuine, P. Cour, and D. D. Rousseau, *Fitting models predicting dates of flowering of temperate-zone trees using simulated annealing*, Plant, Cell & Environment, 21 (1998), pp. 455–466.

[51] M. Colombia, *Colombia, the first country in latin america to have diagnostic tests for the new coronavirus. minsitry of health and social protection 2020. accessed on february 10, 2021.*

[52] M. Colombia, *Ministry of health colombia 2021. accessed on february 20 2021.*

[53] C. N. H. Committee, *Reported cases of 2019-ncov*, tech. rep., Chinese National Health Committee, 2020.

[54] W. Commons, *Map of chile accessed february 7, 2021.*

[55] G. Consolini and M. Materassi, *A stretched logistic equation for pandemic spreading*, Chaos, Solitons & Fractals, 140 (2020), p. 110113.

[56] E. Y. Cramer, E. L. Ray, V. K. Lopez, J. Bracher, A. Brennen, A. J. C. Rivadeneira, A. Gerding, T. Gneiting, K. H. House, Y. Huang, et al., *Evaluation of individual and ensemble probabilistic forecasts of covid-19 mortality in the us*, Medrxiv, (2021).

[57] J. P. Daniels, *Covid-19 cases surge in colombia*, The Lancet, 396 (2020).

[58] G. de Chile, *Censo 2017 chile (accessed february 7, 2021).*

[59] M. de Salud Chile, *Influenza pandemia (h1n1) 2009. circular de vigilancia epidemiología de influenza.*

[60] H. P. de Vladar, *Density-dependence as a size-independent regulatory mechanism*, Journal of Theoretical Biology, 238 (2006), pp. 245–256.

[61] O. Diekmann, H. Heesterbeek, and T. Britton, *Mathematical Tools for Understanding Infectious Disease Dynamics*, Princeton University Press, Dec. 2012.

[62] A. d'Onofrio, *Fractal growth of tumors and other cellular populations: Linking the mechanistic to the phenomenological modeling and vice versa*, Chaos, Solitons & Fractals, 41 (2009), pp. 875–880.

[63] A. d'Onofrio, A. Fasano, and B. Monechi, *A generalization of gompertz law compatible with the gyllenberg–webb theory for tumour growth*, Mathematical Biosciences, 230 (2011), pp. 45–54.

[64] ECDC, *Covid-19 situation update worldwide, as of week 44, updated 11 november 2021. european entre for disease prevention and control (ecdc); november 11, 2021. accessed on november 11, 2021.*

[65] L. Edelstein-Keshet, *Mathematical Models in Biology*, Society for Industrial and Applied Mathematics, Jan. 2005.

[66] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*, Springer Science + Business Media Dordrecht, 1993.

[67] A. Elsevier/North Holland Biomedical Press, ed., *The epidemiology of Ebola hemorrhagic fever in Zaire, 1976*, Elsevier/North Holland Biomedical Press, Amsterdam, 1978.

[68] N. D. Evans, L. J. White, M. J. Chapman, K. R. Godfrey, and M. J. Chappell, *The structural identifiability of the susceptible infected recovered model with seasonal forcing*, Mathematical biosciences, 194 (2005), pp. 175–197.

[69] D. Faranda, I. P. Castillo, O. Hulme, A. Jezequel, J. S. W. Lamb, Y. Sato, and E. L. Thompson, *Asymptotic estimates of SARS-CoV-2 infection counts and their sensitivity to stochastic perturbation*, Chaos: An Interdisciplinary Journal of Nonlinear Science, 30 (2020), p. 051107.

[70] F. H. C. Felix and J. Fontenele, *Instantaneous r calculation for covid-19 epidemic in brazil*, Preprint. medRxiv, (2020).

[71] S. Flaxman et al., *Estimating the effects of non-pharmaceutical interventions on COVID-19 in europe*, Nature, 584 (2020), pp. 257–261.

[72] J. for Colombia, *Coronavirus: what is the impact in colombia? justice for colombia, november 5,2020. accessed on march 5, 2021.*

[73] C. for Disease Control and P. (CDC), *Cdc wonder—aids public information dataset u.s. surveillance (accessed 27 september 2016).*

[74] D. Freire-Flores, N. Llanovarced-Kawles, A. Sanchez-Daza, and Á. Olivera-Nappa, *On the heterogeneous spread of covid-19 in chile*, Chaos, Solitons & Fractals, 150 (2021), p. 111156.

[75] P. J. Garcia et al., The American Journal of Tropical Medicine and Hygiene, 103 (2020), pp. 1765–1772.

[76] T. Gneiting and A. E. Raftery, *Strictly proper scoring rules, prediction, and estimation*, Journal of the American Statistical Association, 102 (2007), pp. 359–378.

[77] B. Gompertz, *On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. in a letter to francis baily, esq. f. r. s. &c. by benjamin gompertz, esq. f. r. s*, Abstracts of the Papers Printed in the Philosophical Transactions of the Royal Society of London, 2 (1833), pp. 252–253.

[78] R. Gutiérrez-jáimez, P. Román, D. Romero, J. J. Serrano, and F. Torres, *A new gompertz-type diffusion process with application to random growth*, Mathematical Biosciences, 208 (2007), pp. 147–165.

[79] I. A. S. R. (IASR), *HIV/AIDS in japan, 2013*, Tech. Rep. 9 (415), National Institute of Infectious Diseases, 2014.

[80] INS, *Covid-19 in colombia, instituto nacional de salud, 2021*.

[81] J. Jiwei, D. Jian, L. Siyu, L. Guidong, L. Jingzhi, D. Ben, W. Guoqing, and Z. Ran, *Modeling the control of covid-19: Impact of policy interventions and meteorological factors*, arXiv e-prints, (2020).

[82] D. S. Jones, ed., *Differential Equations and Mathematical Biology*, Springer Netherlands, 1983.

[83] D. JP, *'everything is collapsing': Colombia battles third covid wave amid unrest. the guardian. june 22, 2021*.

[84] Y.-H. Kao and M. C. Eisenberg, *Practical unidentifiability of a simple vector-borne disease model: Implications for parameter estimation and intervention assessment*, Epidemics, 25 (2018), pp. 89–100.

[85] W. O. Kermack and A. G. McKendrick, *A contribution to the mathematical theory of epidemics*, Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 115 (1927), pp. 700–721.

[86] T. Kirby, *South america prepares for the impact of COVID-19*, The Lancet Respiratory Medicine, 8 (2020), pp. 551–552.

[87] M. Kühleitner, N. Brunner, W.-G. Nowak, K. Renner-Martin, and K. Scheicher, *Best fitting tumor growth models of the von bertalanffy-PütterType*, BMC Cancer, 19 (2019).

[88] M. Kuhn and K. Johnson, *Applied Predictive Modeling*, Springer New York, 2013.

[89] M. Kuhn, K. Johnson, et al., *Applied predictive modeling*, vol. 26, Springer, 2013.

[90] U. Ledzewicz, O. Olumoye, and H. Schättler, *On optimal chemotherapy with a strongly targeted agent for a model of tumor-immune system interactions with generalized logistic growth*, Mathematical Biosciences and Engineering, 10 (2013), pp. 787–802.

[91] Y. Li, H. Campbell, D. Kulkarni, A. Harpur, M. Nundy, X. Wang, and H. Nair, *The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (r) of SARS-CoV-2: a modelling study across 131 countries*, The Lancet Infectious Diseases, 21 (2021), pp. 193–202.

[92] N. Linton, T. Kobayashi, Y. Yang, K. Hayashi, A. Akhmetzhanov, S. mok Jung, B. Yuan, R. Kinoshita, and H. Nishiura, *Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: A statistical analysis of publicly available case data*, Journal of Clinical Medicine, 9 (2020), p. 538.

[93] I. Luz-Sant'Ana, P. Román-Román, and F. Torres-Ruiz, *Modeling oil production and its peak by means of a stochastic diffusion process based on the hubbert curve*, Energy, 133 (2017), pp. 455–470.

[94] J. Ma, J. Dushoff, B. M. Bolker, and D. J. D. Earn, *Estimating initial epidemic growth rates*, Bulletin of Mathematical Biology, 76 (2013), pp. 245–260.

[95] M. Malta, M. V. Vettore, C. M. F. P. da Silva, A. B. Silva, and S. A. Strathdee, *The foreseen loss of the battle against covid-19 in south america: A foretold tragedy*, EClinicalMedicine, 39 (2021).

[96] M. Martcheva, *An Introduction to Mathematical Epidemiology*, Springer US, 2015.

[97] T. McKinley, A. R. Cook, and R. Deardon, *Inference in epidemic models without likelihoods*, The International Journal of Biostatistics, 5 (2009).

[98] C. I. Mendoza, *Inhomogeneous transmission and asynchronic mixing in the spread of covid-19 epidemics*, Frontiers in Physics, 9 (2021).

[99] L. Mercado, *El 1ro de septiembre termina cuarentena y empieza aislamiento selectivo. august 25, 2020. el tiempo.*

[100] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *Equation of state calculations by fast computing machines*, The Journal of Chemical Physics, 21 (1953), pp. 1087–1092.

[101] H. Miao, C. Dykes, L. M. Demeter, J. Cavenaugh, S. Y. Park, A. S. Perelson, and H. Wu, *Modeling and estimation of kinetic parameters and replicative fitness of hiv-1 from flow-cytometry-based growth competition experiments*, Bulletin of mathematical biology, 70 (2008), pp. 1749–1771.

[102] H. Miao, X. Xia, A. S. Perelson, and H. Wu, *On identifiability of nonlinear ode models and applications in viral dynamics*, SIAM review. Society for Industrial and Applied Mathematics, 53 1 (2011), pp. 3–39.

[103] MinCiencia, *Website of the official database for covid-19 research; accessed february 16, 2021 (ministery of science, technology, knowledge and innovation of chile)*.

[104] C. Ministerio de Salud, *Casos confirmados en chile covid-19 (february 15, 20)*.

[105] MinSalud, *Official covid-19 website accessed february 16, 2021 (government of chile)*, 2021.

[106] K. Mizumoto, K. Kagaya, and G. Chowell, *Early epidemiological assessment of the transmission potential and virulence of coronavirus disease 2019 (COVID-19) in wuhan city, china, january–february, 2020*, BMC Medicine, 18 (2020).

[107] K. Mizumoto, K. Kagaya, A. Zarebski, and G. Chowell, *Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the diamond princess cruise ship, yokohama, japan, 2020*, Eurosurveillance, 25 (2020).

[108] A. Morciglio, B. Zhang, G. Chowell, J. M. Hyman, and Y. Jiang, *Mask-ematics: Modeling the effects of masks in covid-19 transmission in high-risk environments*, Epidemiolog, (2021).

[109] C. V. Munayco, A. Tariq, R. Rothenberg, and OTHERS, *Early transmission dynamics of covid-19 in a southern hemisphere setting: Lima-peru: February 29th–march 30th, 2020*, Infectious Disease Modelling, 5 (2020), pp. 338–345.

[110] J. D. Murray, ed., *Mathematical Biology*, Springer New York, 2002.

[111] J.-C. Navarro, J. Arrivillaga-Henríquez, J. Salazar-Loor, and A. J. Rodriguez-Morales, *COVID-19 and dengue, co-epidemics in ecuador and other countries in latin america: Pushing strained health care systems over the edge*, Travel Medicine and Infectious Disease, 37 (2020), p. 101656.

[112] C. M. Newton, *Biomathematics in oncology: Modeling of cellular systems*, Annual Review of Biophysics and Bioengineering, 9 (1980), pp. 541–579.

[113] H. Nishiura and G. Chowell, *The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends*, in Mathematical and Statistical Estimation Approaches in Epidemiology, Springer Netherlands, 2009, pp. 103–121.

[114] H. Nishiura, G. Chowell, H. Heesterbeek, and J. Wallinga, *The ideal reporting interval for an epidemic to objectively interpret the epidemiological time course*, Journal of The Royal Society Interface, 7 (2010), pp. 297–307.

[115] H. Nishiura et al., *Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19)*, International Journal of Infectious Diseases, 94 (2020), pp. 154–155.

[116] A. Ohnishi, Y. Namekawa, and T. Fukui, *Universality in COVID-19 spread in view of the gompertz function*, Progress of Theoretical and Experimental Physics, 2020 (2020).

[117] S. Ohnishi, T. Yamakawa, and T. Akamine, *On the analytical solution for the pütter–bertalanffy growth equation*, Journal of Theoretical Biology, 343 (2014), pp. 174–177.

[118] E. operational readiness and preparedness, *2015 ebola response roadmap—situation report—14 october 2015 (accessed 17 october 2015)*, tech. rep., World health organization, 2015.

[119] B. Pell, Y. Kuang, C. Viboud, and G. Chowell, *Using phenomenological models for forecasting the 2015 ebola challenge*, Epidemics, 22 (2018), pp. 62–70.

[120] L. Peng, W. Yang, D. Zhang, C. Zhuge, and L. Hong, *Epidemic analysis of COVID-19 in china by dynamical modeling*, medRxiv, (2020).

[121] A. Physikat, *Uddrag af det kongelige sundhedskollegium aarsberetning for 1853 (in danish; accessed 27 september 2016)*, tech. rep., docplayer, 1853.

[122] Portafolio, *The health emergency in colombia will last three more months. portafolio. 2021 february 25.*

[123] T. A. Press, *Colombia reaches 1 million confirmed coronavirus cases, 24 october 2020. abc news.*

[124] B. Radio, *Covid-19 outbreak: 74 older adults were infected with coronavirus in manizales. blu radio news, august 20, 2020. accessed on march 4, 2021.*

[125] J. M. Read, J. R. Bridgen, D. A. Cummings, A. Ho, and C. P. Jewell, *Novel coronavirus 2019-ncov: early estimation of epidemiological parameters and epidemic predictions*, medRxiv, (2020).

[126] M. Renardy, D. Kirschner, and M. Eisenberg, *Structural identifiability analysis of age-structured pde epidemic models*, Journal of Mathematical Biology, 84 (2022), pp. 1–30.

[127] F. J. Richards, *A flexible growth function for empirical use*, Journal of Experimental Botany, 10 (1959), pp. 290–301.

[128] P. Román-Román, D. Romero, M. Rubio, and F. Torres-Ruiz, *Estimating the parameters of a gompertz-type diffusion process by means of simulated annealing*, Applied Mathematics and Computation, 218 (2012), pp. 5121–5131.

[129] P. Román-Román, D. Romero, and F. Torres-Ruiz, *A diffusion process to model generalized von bertalanffy growth patterns: Fitting to real data*, Journal of Theoretical Biology, 263 (2010), pp. 59–69.

[130] P. Román-Román, J. Serrano-Pérez, and F. Torres-Ruiz, *Some notes about inference for the lognormal diffusion process with exogenous factors*, Mathematics, 6 (2018), p. 85.

[131] P. Román-Román and F. Torres-Ruiz, *Modelling logistic growth by a new diffusion process: Application to biological systems*, Biosystems, 110 (2012), pp. 9–21.

[132] P. Román-Román and F. Torres-Ruiz, *The nonhomogeneous lognormal diffusion process as a general process to model particular types of growth patterns*, Lecture Notes of Seminario Interdisciplinare di Matematica, 12 (2015), pp. 201–219.

[133] P. Román-Román and F. Torres-Ruiz, *A stochastic model related to the richards-type growth curve. estimation by means of simulated annealing and variable neighborhood search*, Applied Mathematics and Computation, 266 (2015), pp. 579–598.

[134] K. Roosa and G. Chowell, *Assessing parameter identifiability in compartmental dynamic models using a computational approach: application to infectious disease transmission models*, Theoretical Biology and Medical Modelling, 16 (2019).

[135] K. Roosa, Y. Lee, R. Luo, A. Kirpich, R. Rothenberg, J. Hyman, P. Yan, and G. Chowell, *Real-time forecasts of the COVID-19 epidemic in china from february 5th to february 24th, 2020*, Infectious Disease Modelling, 5 (2020), pp. 256–263.

[136] K. Roosa, R. Luo, and G. Chowell, *Comparative assessment of parameter estimation methods in the presence of overdispersion: a simulation study*, Math Biosci Eng, 16 (2019), pp. 4299–313.

[137] S. Roy, R. Dutta, and P. Ghosh, *Towards dynamic lockdown strategies controlling pandemic spread under healthcare resource budget*, Applied Network Science, 6 (2021).

[138] M. S., *Are covid vaccination programmes working? scientists seek first clues*, Nature, (2021).

[139] T. Sauer, T. Berry, D. Ebeigbe, M. M. Norton, A. J. Whalen, and S. J. Schiff, *Identifiability of infection model parameters early in an epidemic*, SIAM Journal on Control and Optimization, (2021), pp. S27–S48.

[140] L. Schüler, J. M. Calabrese, and S. Attinger, *Data driven high resolution modeling and spatial analyses of the covid-19 pandemic in germany*, medRxiv, (2021).

[141] L. A. Segel and L. Edelstein-Keshet, *A Primer on Mathematical Models in Biology*, Society for Industrial and Applied Mathematics, Mar. 2013.

[142] D. W. Shanafelt, G. Jones, M. Lima, C. Perrings, and G. Chowell, *Forecasting the 2001 foot-and-mouth disease epidemic in the uk*, EcoHealth, 15 (2018), pp. 338–347.

[143] J. M. Shultz, R. C. Berg, and OTHERS, *Complex correlates of colombia's covid-19 surge*, The Lancet Regional Health - Americas, 3 (2021), p. 100072.

[144] A. Sommer, *The 1972 smallpox outbreak in khulna municipality, bangladesh: II. effectiveness of surveillance and containment in urban epidemic control*, American Journal of Epidemiology, 99 (1974), pp. 303–313.

[145] S. H. Strogatz, *Nonlinear Dynamics and Chaoss with Applications to Physics, Biology, Chemistry, and Engieening*, Perseus Books, Reading, MA, 1994.

[146] B. T., *Covid-19 in latin america*, The Lancet. Infectious diseases, (2020).

[147] A. Talkington and R. Durrett, *Estimating tumor growth rates in vivo*, Bulletin of Mathematical Biology, 77 (2015), pp. 1934–1954.

[148] G. Tapiwa et al., *Estimating the generation interval for coronavirus disease (covid-19) based on symptom onset data, march 2020*, Euro Surveill., 25 (2020).

[149] A. Tariq, J. M. Banda, P. Skums, et al., *Transmission dynamics and forecasts of the covid-19 pandemic in mexico, march-december 2020*, PLOS ONE, 16 (2021), pp. 1–34.

[150] A. Tariq, T. Chakhaia, S. Dahal, A. Ewing, X. Hua, S. K. Ofori, O. Prince, A. D. Salindri, A. E. Adeniyi, J. M. Banda, P. Skums, R. Luo, L. Y. Lara-Díaz, R. Bürger, I. C.-H. Fung, E. Shim, A. Kirpich, A. Srivastava, and G. Chowell, *An investigation of spatial-temporal patterns and predictions of the coronavirus 2019 pandemic in colombia, 2020–2021*, PLOS Neglected Tropical Diseases, 16 (2022), pp. 1–33.

[151] A. Tariq et al., *Real-time monitoring the transmission potential of covid-19 in singapore, march 2020*, BMC Medicine, (2020).

[152] A. Tariq, K. Roosa, K. Mizumoto, and G. Chowell, *Assessing reporting delays and the effective reproduction number: The ebola epidemic in drc, may 2018–january 2019*, Epidemics, 26 (2019), pp. 128–133.

[153] 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, 15 (2021), p. e0009070.

[154] T. Team, *The epidemiological characteristics of an outbreak of 2019 novelcoronavirus diseases (covid-19) −china, 2020.*

[155] E. Tiempo, *Government defined date of start of vaccination and purchase of more doses. el tiempo, january 29, 2021.*

[156] E. Tjørve and K. M. Tjørve, *A unified approach to the richards-model family for use in growth analyses: Why we need only two model forms*, Journal of Theoretical Biology, 267 (2010), pp. 417–425.

[157] K. M. Tjørve and E. Tjørve, *A proposed family of unified models for sigmoidal growth*, Ecological Modelling, 359 (2017), pp. 117–127.

[158] K. M. C. Tjørve and E. Tjørve, *The use of gompertz models in growth analyses, and new gompertz-model approach: An addition to the unified-richards family*, PLOS ONE, 12 (2017), p. e0178691.

[159] R. Todeschini, V. Consonni, and M. Pavan, *A distance measure between models: a tool for similarity/diversity analysis of model populations*, Chemometrics and Intelligent Laboratory Systems, 70 (2004), pp. 55–61.

[160] B. M. Trejos-Herrera AM, Vinaccia S, *Coronavirus in colombia: Stigma and quarantine.*, Journal of global health, 10 (2020).

[161] A. Tsoularis and J. Wallace, *Analysis of logistic growth models*, Mathematical Biosciences, 179 (2002), pp. 21–55.

[162] N. Tuncer and T. T. Le, *Structural and practical identifiability analysis of outbreak models*, Mathematical biosciences, 299 (2018), pp. 1–18.

[163] M. E. Turner, E. L. Bradley, K. A. Kirk, and K. M. Pruitt, *A theory of growth*, Mathematical Biosciences, 29 (1976), pp. 367–373.

[164] E. A. Undurraga, G. Chowell, and K. Mizumoto, *COVID-19 case fatality risk by age and gender in a high testing setting in latin america: Chile, march–august 2020*, Infectious Diseases of Poverty, 10 (2021).

[165] S. Y. D. Valle et al., *Summary results of the 2014-2015 DARPA chikungunya challenge*, BMC Infectious Diseases, 18 (2018).

[166] P. van den Driessche and J. Watmough, *Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission*, Mathematical Biosciences, 180 (2002), pp. 29–48.

[167] A. R. Varela, L. J. H. Florez, G. Tamayo-Cabeza, and OTHERS, *Factors associated with sars-cov-2 infection in bogotá, colombia: Results from a large epidemiological surveillance study*, Lancet Regional Health. Americas, 2 (2021), p. 100048.

[168] G. L. Vasconcelos, A. M. Macêdo, R. Ospina, F. A. Almeida, G. C. Duarte-Filho, A. A. Brum, and I. C. Souza, *Modelling fatality curves of COVID-19 and the effectiveness of intervention strategies*, PeerJ, 8 (2020), p. e9421.

[169] G. L. Vasconcelos, A. M. S. Macêdo, G. C. Duarte-Filho, A. A. Brum, R. Ospina, and F. A. G. Almeida, *Power law behaviour in the saturation regime of fatality curves of the COVID-19 pandemic*, medRxiv, (2020).

[170] C. Viboud, L. Simonsen, and G. Chowell, *A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks*, Epidemics, 15 (2016), pp. 27–37.

[171] C. Viboud, K. Sun, R. Gaffey, M. Ajelli, L. Fumanelli, S. Merler, Q. Zhang, G. Chowell, L. Simonsen, and A. Vespignani, *The RAPIDD ebola forecasting challenge: Synthesis and lessons learnt*, Epidemics, 22 (2018), pp. 13–21.

[172] R. V. V. Vidal, ed., *Applied Simulated Annealing*, Springer Berlin Heidelberg, 1993.

[173] M. Vogels, R. Zoeckler, D. M. Stasiw, and L. C. Cerny, *P. f. verhulst's "notice sur la loi que la populations suit dans son accroissement" from correspondence mathematique et physique. ghent, vol. x, 1838*, Journal of Biological Physics, 3 (1975), pp. 183–192.

[174] L. von Bertalanffy, *A quantitative theory of organic growth (inquiries on growth laws. ii)*, Human Biology a record of research, 10 (1938), pp. 181–213.

[175] L. von Bertalanffy, *Quantitative laws in metabolism and growth*, The Quarterly Review of Biology, 32 (1957), pp. 217–231. PMID: 13485376.

[176] E. Vynnycky and R. White, *An Introduction to Infectious Disease Modelling*, Oxford University Press, 2010.

[177] P. G. T. Walker et al., *The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries*, Science, 369 (2020), pp. 413–422.

[178] J. Wallinga and M. Lipsitch, *How generation intervals shape the relationship between growth rates and reproductive numbers*, Proceedings of the Royal Society B: Biological Sciences, 274 (2006), pp. 599–604.

[179] X.-S. Wang, J. Wu, and Y. Yang, *Richards model revisited: Validation by and application to infection dynamics*, Journal of Theoretical Biology, 313 (2012), pp. 12–19.

[180] ——, *Richards model revisited: Validation by and application to infection dynamics*, Journal of Theoretical Biology, 313 (2012), pp. 12–19.

[181] T. E. Wheldon, *Mathematical models in cancer research*, Hilger, Bristol and Philadelphia 1988, 1988.

[182] WHO, *Outbreak of ebola hemorrhagic fever, uganda, august 2000–january 2001*, tech. rep., World health organization (WHO), 2001.

[183] ——, *Yellow fever situation reports, angola, situation reports march 2016–july 2016 (accessed 20 january 2019)*, tech. rep., World health organization (WHO), 2016.

[184] ——, *Plague outbreak situation reports, madagascar, october 2017–december 2017 (accessed 20 january 2019)*, tech. rep., World health organization (WHO), 2017.

[185] ——, *Who director-general's opening remarks at the media briefing on covid-19-11 march 2020*, tech. rep., World Health Organization (WHO), 2020.

[186] O. world in data, *Total covid-19 tests per 1,000 people: Our world in data, 2020. accessed on september 24, 2021.*

[187] H. Wu, H. Zhu, H. Miao, and A. S. Perelson, *Parameter identifiability and estimation of hiv/aids dynamic models*, Bulletin of mathematical biology, 70 (2008), pp. 785–799.

[188] P. Yan and G. Chowell, *Quantitative Methods for Investigating Infectious Disease Outbreaks*, Springer International Publishing, 2019.

[189] C. You et al., *Estimation of the time-varying reproduction number of COVID-19 outbreak in china*, International Journal of Hygiene and Environmental Health, 228 (2020), p. 113555.

[190] C. Zhan, Y. Zheng, Z. Lai, T. Hao, and B. Li, *Identifying epidemic spreading dynamics of COVID-19 by pseudocoevolutionary simulated annealing optimizers*, Neural Computing and Applications, 33 (2020), pp. 4915–4928.

[191] S. Zhao, S. S. Musa, H. Fu, D. He, and J. Qin, *Simple framework for real-time forecast in a data-limited situation: the zika virus (ZIKV) outbreaks in brazil from 2015 to 2016 as an example*, Parasites & Vectors, 12 (2019).

[192] Z. Zhao, X. Li, F. Liu, G. Zhu, C. Ma, and L. Wang, *Prediction of the COVID-19 spread in african countries and implications for prevention and control: A case study in south africa, egypt, algeria, nigeria, senegal and kenya*, Science of The Total Environment, 729 (2020), p. 138959.

[193] C. Zhou, *Evaluating new evidence in the early dynamics of the novel coronavirus COVID-19 outbreak in wuhan, china with real time domestic traffic and potential asymptomatic transmissions*, medRxiv, (2020).

[194] L. Zou et al., *SARS-CoV-2 viral load in upper respiratory specimens of infected patients*, New England Journal of Medicine, 382 (2020), pp. 1177–1179.