Mathematical and statistical modeling of epidemic data : application to the novel COVID-19 outbreak

Kayode Oshinubi

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : ISCE - Ingénierie pour la Santé la Cognition et l'Environnement
Spécialité : MBS - Modèles, méthodes et algorithmes en biologie, santé et environnement
Unité de recherche : AUTONOMIE, GÉRONTOLOGIE, E-SANTÉ, IMAGERIE & SOCIÉTÉ

Modélisation mathématique et statistique des données épidémiques : application à la nouvelle épidémie de COVID-19

Mathematical and statistical modeling of epidemic data: application to the novel COVID-19 outbreak

Présentée par :

Kayode OSHINUBI

Direction de thèse :

Mustapha RACHDI Université Grenoble Alpes	Directeur de thèse
Jacques DEMONGEOT émérite, UGA	Co-directeur de thèse

Rapporteurs :

SOPHIE DABO-NIANG
Professeur des Universités, UNIVERSITE DE LILLE
HALIMA BENSMAIL
Professeur associé, Hamad Bin Khalifa University

Thèse soutenue publiquement le **26 octobre 2022**, devant le jury composé de :

MUSTAPHA RACHDI Professeur des Universités, UNIVERSITE GRENOBLE ALPES	Directeur de thèse
SOPHIE DABO-NIANG Professeur des Universités, UNIVERSITE DE LILLE	Rapporteure
HALIMA BENSMAIL Professeur associé, Hamad Bin Khalifa University	Rapporteure
JACQUES DEMONGEOT Professeur des Universités émérite, UNIVERSITE GRENOBLE ALPES	Co-directeur de thèse
JOSE LABARERE Professeur des Univ. - Praticien hosp., UNIVERSITE GRENOBLE ALPES	Président
PIERRE MAGAL Professeur des Universités, UNIVERSITE DE BORDEAUX	Examinateur

Invités :

OLUMIDE LONGE
Professeur, Academic City University College, Accra, Ghana

Abstract

This thesis is devoted to the mathematical and statistical modeling of epidemic data and It is divided into two broad parts, which are subdivided into different sections. The modeling of infectious diseases has been a subject of interest to researchers, policy makers, and medical practitioners, most especially during the recent global COVID-19 pandemic, which It has been devastating to the health infrastructure and socio-economic status of many nations. It has affected mobility and interaction among citizens due to the many daily new cases and deaths. Hence, the need to contribute to understanding the mechanisms of virulence and spread using different mathematical and statistical modeling approaches. The first part is dedicated to the mathematical modeling aspect, which consists of the deterministic and discrete approaches to epidemiology modeling, which in this case is mainly focused on the COVID-19 pandemic. The daily reproduction number of the COVID-19 outbreak calculation is approached by discretization using the idea of deconvolution and a unique biphasic pattern is observed that is more prevalent during the contagiousness period across various countries. Furthermore, a discrete model is formulated from Usher's model in order to calculate the life span loss due to COVID-19 disease and to also explain the role of comorbidities, which are very essential in the disease spread and its dynamics at an individual level. Also, the formulation of Susceptible-Infectious-Geneanewsusceptible-Recovered (SIGR) age-dependent modeling is proposed in order to perform some mathematical analysis and present the role of different epidemiology parameters, most especially vaccination, and finally, a new technique to identify the point of inflection on the smoothed curves of the new infected pandemic cases using the Bernoulli equation is presented. This procedure is important because not all countries have reached the turning point (maximum number of daily cases) in the epidemic curve. The approach is used to calculate the transmission rate and the maximum reproduction number for various countries. The statistical modeling of the COVID-19 pandemic using various data analysis models (namely machine and deep learning models) is presented in the second part in order to understand the dynamics of the pandemic in different countries and also predict and forecast the daily new cases and deaths due to the disease alongside some socio-economic parameters. It is observed that the prediction and forecasting are consistent with the disease evolution at different waves in these countries and that there are socio-economic determinants of the disease depending on whether the country is developed or developing. Also, the study of the shapes and peaks of the COVID-19 disease is presented. The peaks of the curves of the daily new cases and deaths are identified using the spectral analysis method, which enables the weekly peak patterns to be visible. Finally, the clustering of different regions in France due to the spread of the disease is modeled using functional data analysis. The study shows clear differences between the periods when vaccination has not been introduced (but only non-pharmaceutical mitigation measures) and when it was introduced. The results presented in this thesis are useful to better understand the modeling of a viral disease, the COVID-19 virus.

Résumé

Cette thèse est consacrée à la modélisation des épidémies et elle est divisée en deux parties qui se subdivisent elles-mêmes en différentes sections. La modélisation des maladies infectieuses a suscité depuis de nombreuses années (la première modélisation de la propagation de la variole de D. Bernoulli date de 1760)) l'intérêt des décideurs et des médecins, plus particulièrement lors de la récente pandémie de COVID-19, qui a déstabilisé les infrastructures de santé, altéré le statut socio-économique et affecté les interactions entre citoyens dans de nombreux pays, en raison de multiples cas et décès, d'où la nécessité de comprendre les mécanismes de virulence et de propagation, à l'aide de diverses approches de modélisation mathématique et statistique. La première partie est consacrée à l'aspect modélisation mathématique d'une épidémie, qui consiste en des approches déterministes et discrètes de la propagation épidémique, principalement axée sur la pandémie de COVID-19. Le nombre quotidien de reproduction durant la période de contagiosité est approché par discrétisation, fondée sur l'idée de dé-convolution, et la forme bi-phasique a été la plus souvent calculée dans divers pays, ceci étant possible, si on suppose que la maladie a commencé par un unique patient infectieux « zéro » au démarrage de l'épidémie et que les taux de transmission et la taille de la population susceptible sont constants en phase de croissance exponentielle des nouveaux cas observés. Ensuite, un modèle discret été formulé à partir du modèle d'Usher, afin de calculer la perte de durée de vie des infectés, due à la maladie COVID-19 et afin d'expliquer également le rôle des comorbidités, qui est très essentiel dans la propagation de la maladie et sa dynamique au niveau individuel. En outre, une formulation de la modélisation de la dynamique épidémique dépendante de l'âge des infectés, de type Susceptible-Infectious-Geneanewsusceptible-Recovered (SIGR), a été proposée, afin d'analyser le rôle de différents paramètres épidémiologiques, plus particulièrement celui de la vaccination. Enfin, une nouvelle technique a été proposée pour identifier le point d'inflexion sur les courbes lissées des nouveaux cas infectés, cela à l'aide de l'équation princeps de Bernoulli. Cette procédure est importante, car les vagues de nombreux pays n'ont pas atteint le point de retournement (maximum de cas quotidiens) de leur courbe épidémique, en particulier pour des vagues proches, dans lesquelles on peut avoir une succession d'épaulements, mais pas de maximas. La modélisation statistique de la pandémie de COVID-19, à l'aide de divers modèles d'apprentissage automatique et profond, est présentée dans la deuxième partie, afin de comprendre la dynamique de la pandémie dans différents pays, ainsi que de prédire et de prévoir les nouveaux cas quotidiens et les décès dus à la maladie, ainsi que certains paramètres socio-économiques. On observe que la prédiction et la prévision sont cohérentes avec l'évolution de la maladie durant différentes vagues dans ces différents pays, et qu'il existe des déterminants socio-économiques de la maladie, selon que le pays est développé ou en voie de développement. De plus, l'étude du nombre et des formes des pics de la maladie COVID-19 est présentée. Les pics des courbes des nouveaux cas et des décès quotidiens sont identifiés à l'aide d'une méthode d'analyse spectrale, qui permet de visualiser les modèles de pics hebdomadaires. Enfin, la classification des départements

français vis-à-vis de la propagation de la maladie est effectuée par analyse fonctionnelle des données. Cette dernière étude montre des différences nettes entre la période où la vaccination n'a pas encore été introduite (seulement des mesures d'atténuation non pharmaceutiques) et celle où elle l'a été. Les résultats de cette thèse seront particulièrement utiles dans le futur pour mieux comprendre les mécanismes de propagation d'une maladie virale, ici la maladie due au virus SARS-CoV-2.

Dedication

The thesis is dedicated to God Almighty, the Giver of Knowledge. To my mother - Mrs B. Oshinubi, who has been a source of inspiration despite her lack of a university education.

Acknowledgements

I thank God for assisting me in starting this journey and seeing it through to completion because without him, I would not have been able to complete this doctoral journey.
Prof. Mustapha Rachdi and Prof. Jacques Demongeot, my wonderful supervisors, deserve my heartfelt thanks. Your help has been invaluable throughout this journey. Despite meeting you for supervision during the most difficult period of my life, you made the journey interesting for me. I truly appreciate it. I am grateful to the members of my doctoral monitoring committee, Prof. José Labarere and Prof. Marc Hommel, for their various contributions, encouragement, and support throughout this research work.
I also appreciate the support giving to me by my doctoral school. Thank you Prof. Chris Moulin and Orélie Garzena. I am also indeed thankful to Prof. Bernard Tourancheau for his support during the most challenging time of my doctoral journey. Thank you for everything, Prof. Roseline Ogunrinde and Prof. Olumide Longe. These are the people I go to when I need to get my bearings on this journey. Thank you for your encouraging words.
During my time in Grenoble, I was able to make friends both inside and outside of our laboratory, at my church, the International Church of Grenoble, and throughout the city of Grenoble. In one way or another, these people have made my doctoral journey interesting and lively, and I am grateful to you all. You guys are the best, Africa community, especially the Nigerian community, with whom we occasionally hang out to discuss and have some fun. Also, I want to express my gratitude to my wonderful close friends who are always there for me and pray for me when things get difficult. Thank you, Joshua, Rose, Tosin, Freeman, Toyin, Moshood, Bayo, Seun, Dr. Muideen, Augustina, and Dr. Kunle, to name a few. I appreciate everyone. Dr. Muyiwa deserves special recognition for igniting my interest in infectious disease modeling and for his numerous collaborations over the years.
Without the support of my sponsors, Petroluem Technology Development Fund (PTDF) and Campus France Africa Unit, I would not have been able to finance my doctoral studies. They provided all necessary assistance so that I could survive without financial stress. Thank you.
Finally, thanks to my family, especially to my mother, for all of her help in so many ways. Thank you to my father, sister, and brothers for always being there for me.

Contents

Chapter 1

Introduction

1.1 General Background

Epidemiology and infectious disease modeling and its application to actual viral disease data has been envisioned by many researchers over the years. The modeling of infectious diseases started as early as 1760 by Daniel Bernoulli [1] who proposed the famous deterministic mathematical model Susceptible-Infected (SI) to solve the epidemic wave of smallpox, but it is a known fact that the stages of infectious diseases are the stages endemic, epidemic and eradicated while the model he presented defined but not modeled the endemic state.
He said:
"Le retour d'une épidémie longtemps suspendue fait un ravage plus terrible dans une seule année qu'une endémie uniforme ne pourrait faire pendant un nombre d'années considérable.", which means in english:
"The return of a long-suspended epidemic wreaks more terrible havoc in a single year than a uniform endemic could do for a considerable number of years."
This shortcoming has now been taken into account because we have had many models of diseases whose endemic state has been considered due to their endemicity. The improvement of the SI model is now extended to models such as Susceptible-Infected-Recovered (SIR), Susceptible-Exposed-Infected-Recovered (SEIR), Vaccinated-Susceptible-Infected-Recovered-Deaths (VSIRD), Susceptible-Infected-Recovered-Deaths (SIRD), etc. which provide an in-depth understanding of the dynamics of viral diseases. Figure 1.1 is a simple SEIR model where α is the contact rate and γ is the recovery rate.

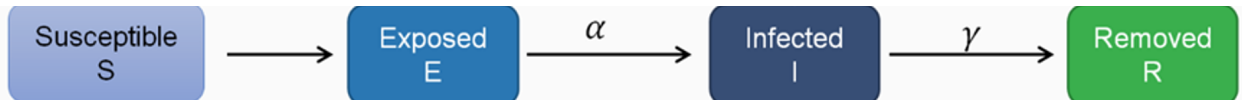


Figure 1.1: Schematic diagram of a simple SEIR model.

The problem of the initial conditions of an epidemy has been already evoked by Bernoulli [1], but in general it is at best only through the hypothesis of a unique "patient zero" allowing the definition on him of his initial daily reproduction number [2]. The patient zero, also called index or primary case is the first patient identified in a given population during an epidemiological investigation. Despite the considerable means of current investigation, in particular those of the World Health Organization (WHO) and the member governments of the WHO,

this patient is rarely identified, despite some cases as the H1N1 epidemic in Mexico [3]. In general, the sources of an infectious disease spread in a given reservoir indicates with their subsequent clusters the possible existence of an outbreak, but this search is in general very long and difficult as for HIV in North America [4].
Concerning these initial clusters, from observations made during investigations of the start of the outbreak in some countries [5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19], it is sometime possible to get spatial and temporal information on the start of the epidemy, but rarely these studies allows estimating the parameters $S(0)$ and $\tau(0)$ in the concerned population and worse, they give no indication of how long they remain constant. This is the reason why we assumed that they remained constant only during the period of exponential growth of new cases observed. A way to solve this problem could be to use the probability distribution given in some of these articles for the length r of the contagiousness period (in general a not-symmetrical lognormal distribution) considered as a random variable and then, for a sample of values of r, simulate the case by adding another noise on the new observed cases X_i at day i of the pandemic, e.g., a Gaussian noise concentrated on the mean values given by different databases devoted to the new cases counting.
The world is recently facing with a new viral disease named COVID-19, whose modeling is the main focus of this thesis. The virus was first discovered in Wuhan China in December, 2019 and since then it has spread to more than 230 countries. There are different symptoms the host of this highly contagious virus do experience which vary in different ways, they are fever, shortness of breath, fatigue, coughing, loss of smell and taste, difficult in breathing and other lower respiratory effects. It is also observed that we have more asymptomatic than those who are symptomatic and it takes about 5-6 days on the average for symptoms to show, and however, it can take up to 14 days in some cases.
As published by World Health Organization (WHO) [20] in March 2020, there are four types of COVID-19 transmission: no cases or no reported cases for at least 28 days, sporadic cases (one or more cases, imported or locally acquired, detected in the last 14 days), clusters of cases (most cases of local transmission linked in chains of transmission detected in the last 14 days are predominantly limited to well-defined clusters), and community transmission (one or more cases, imported or locally acquired, detected in the last 14 days). The number of people infected per unit time is referred to as the incidence in epidemiology. This incidence increases as the disease spreads from persons to persons.
The spread of the disease from patient zero to other susceptible part of the population during the contagiousness period without public health interventions is measured by the basic reproduction rate. This parameter is important in epidemiology modeling and it will be critically discussed and estimated in this thesis.
This spread is more severe in people with comorbidities (kidney failure, obesity, cancer, complicated hypertension, chronic liver disease, psychiatric disorders, dementia, respiratory and heart failure) and severely immunocompromised persons (people with rare and serious pre-existing conditions or severe disabilities, history of stroke, patients with autoimmune diseases or aggressive immunosuppressive therapy, dialysis patients, and hematopoietic stem cell, solid organ or recent bone marrow transplant recipients) due to their high risk factor and vulnerability to infectious diseases.
It has been on the top of the agenda of various stakeholders like researchers, scientists, politicians and medical practitioners to find a way to solve this global menace which has cripple health facilities and economy globally due to its high rate of incidence and fatalities of the disease which make it to be a major concern and significant disruption has been caused to the

socio-economic structure. At the beginning of the pandemic, delays in reporting COVID-19 cases and incubation period (the number of days before an infected individual starts showing symptoms) affect policy making such as the introduction of mitigation measures. Because many countries most especially developing countries were not prepared even though it appears no countries was prepared for it so, there were many efforts to scale up Intensive Care Unit (ICU) beds, ventilators, awareness on sneezing and coughing in elbow, discouraging large gatherings, setting up new isolation centers, encouraging awareness on washing of hands, provision and use of sanitizers and masks, as well as various medical interventions and equipment necessary to help those who are critically ill depending on their symptoms and age because their was a rapid increase in the number of hospitalization cases. Also, place of work and schools started running their activities online in order to encourage e-learning and teleworking so as to avoid physical contacts and interactions like handshakes and greetings with hugs and kisses, and due to isolation, many were affected psychologically with little or no support from stakeholders. Till date, all efforts to mitigate its spread which are ensuring facilities that are open put hygiene concepts (well ventilated environment, sanitizers and nose masks) in place, lock-downs (partial or total), curfew, face-masks, limit in the number of people gathering, restriction of non-essential businesses, stoppage of social, recreation and sporting activities, aggressive contact tracing, massive testing and its prerequisite for travels, social distancing, and administration of different doses of vaccine to the population especially the most vulnerable (elderly) has not yield major result and we have seen many reinfections even in fully vaccinated people but we can't deny the fact that it has helped to reduce hospitalization and severe illness because the virus keep mutating most especially the Delta and Omicron variants which are known for their high contagiousness.

There are a lot of concerns when this outbreak will be endemic because it can't continue to cripple activities all over the world because mobility, businesses and tourism must continue and if possible to live with it as normal flu however, the aggressive vaccination campaign its still ongoing whereby more doses is prescribed by medical scientist so as to proper manage the situation we have found ourselves. We have also observed that there are some demographic determinants of the spread of the disease, since we do have more incidence most especially when lock-downs are eased, during festive and winter period and less cases in other season. The spread and fatality also depends on age (severe in elderly and less severe in young and children) but due to its mutation, the recent Omicron variant is highly contagious, then the epidemic curve shows that children and adult closer to children (like teachers and parents) are more infected, reason why there is a call for vaccination of children which was not the case at the beginning of the pandemic.

Developing an infectious disease model appropriate for the current COVID-19 epidemic that can predict the epidemic's development trend and can also fit an epidemic that has already occurred hence, our goal in this thesis is to contribute to the modeling of COVID-19 outbreak in two directions namely (i) the mathematical modeling and (ii) statistical modeling. In the next section, we present an overview of some literature work in these two directions.

1.2 Systematic Review of Literature

People across the globe are looking for a way to put an end to the COVID-19 pandemic even though on medical level, scientists are currently working on drugs to combat the disease spread and there were different approaches and suggestions by many researchers, some of these contributions in literature will be discussed in this section based on the direction of

this thesis in terms of mathematical modeling, statistical modeling, public health policies and there medical implications.
A lot of mathematical and statistical models have been proposed by researchers in order to suggest best way to combat the spread of the disease and to understand the disease population dynamics. On searching the key word 'Mathematical and Statistical modeling of COVID-19' on cross-reference, there are thousands of articles on this subject which only few of them will be discussed and the summary of models that have been developed and applied to the study of COVID-19 outbreak ranges from mathematical to statistical and theoretical ones, these models employing insight from mathematics, physics, statistics, medicine, epidemiology and infectious disease.
The models summarized in this section are of the forms: difference equation, ordinary differential equation, partial differential equation, matrix model, numerical model, individual-based model, population-based model, regression model, survival function, hazard function, stochastic process, and deep learning model. Discrete-time dynamics are used for stochastic and deterministic population growth models while continuous-time dynamics can be used for host-virus models, stochastic population growth models and advection-diffusion models. Also, spatially homogeneous, deterministic and delay formulations are mostly used for advection-diffusion models, host-virus models and population growth models. Age-structured population models are characterized by linear system of difference equations and can also include stochastic effect and agent-based models that simulate actions and interactions among individuals within the population. They are computer model and can be deterministic or stochastic. Numerical models solve complex dynamical systems and realistic models. Diffusion equation can be used to study the evolution of these models. The idealistic assumption of semi-mechanistic model is that the country is viewed as a closed environment, with all infections assumed to occur within the population. Lastly, statistical tools help to formulate models based on artificial intelligence, survival analysis, machine learning, logistic regression, random effects and sometimes these models include spatial effects.
The pandemic started at a time when many countries were not prepared for the impact the spread would have on the daily activities of the people and scientists are trying to understand if this strain will be like other Coronaviruses strains. Since the pandemic started in China, a model to study the country preparedness and management was proposed in [21]. In [22] authors used logistic growth model and SIR model to study the outbreak in India which is one of the country that was badly hit by the pandemic and it is a country with a very high population so, there was a lot of interactions among people and movement of people from one place to another enabling contacts which lead to high rate of the spread of the disease [23]. There are different waves or exponential phases of the pandemic, and most countries are in their fourth wave while others are already in their fifth wave. A non-linear mathematical model was proposed in [24] to forecast different waves of COVID-19 pandemic. Also, authors in [25] proposed a model for the multiple epidemic waves to study different exponential growth phases of the pandemic. In [26] authors formulated a SEIR compartment population dynamics model with an added formite term. The formite term is analyzed as a means of statistically quantifying the contribution of contaminated fomites to the spread of COVID-19 pandemic. The model helps to guide health policy makers to know the contribution of some fomites to the spread of infection in COVID-19 virus. The use of artificial intelligence techniques is important in infectious disease modeling. In [27], authors developed an artificial intelligence tool which combines epidemiology model and medical data. This approach helped to understand Spanish population movements based on mobile phone

geographically located information to determine their economy activities hence, an economic impact is predicted alongside spatio-temporal spread of COVID-19. Travel restrictions is one of the measures implemented which affected many countries whose national income depends heavily on tourism. [28] presented and discussed mathematical and statistical analysis of this measures and its effects in tourism sector. The author in [29] simulated COVID-19 cases in Nigeria to better understand the peaks in the evolution of the disease recorded using SIR model with certain parameters to be optimized by using the Nelder-Mead algorithm with maximum likelihood estimator.

The use of fractional derivative to model physical problems has played important role in epidemics modeling. It is known to understand the physical meaning of real data fielded into the model even though the model is old, and there has been some improvements in recent time. Some of those improvements include applying it to COVID-19 outbreak modeling. In [30] authors proposed a fuzzy fractional model for COVID-19 epidemic and due to high transmission rate of the disease, in [31], fractional order of SIR model was used to understand and predict the transmission of COVID-19 disease.

Age dependent modeling helps to better understand the dynamics of the disease at sub-population level. Some articles have been devoted to study the role of vaccination at sub-population level and the spread of the disease based on age whose fatality rate is higher in elderly class but for the young class, the transmission rate is high. Age structured compartment model was proposed to study the spread of COVID-19 in Belgium [32] and in Switzerland [33]. To see the effect of vaccination in age modeling, authors in [34] used available data at national level in Italy to validate their proposed model.

The effect of the pandemic on mental health of the people cannot be overemphasized as many fell into depression and anxiety of the unknown and of what will happen next most especially health workers, teachers, workers, students and lecturers due to isolation when various control measures were put in place. As a result, many were separated from families and loved ones, people could not interact because of social distancing, workers were forced to do teleworking and students to start taking classes online to reduce physical interactions. There were reports that psychiatric patients at home were increasing because there was little or no support on mental health during the lockdown. Some studies in [35, 36] focus on mental health during the COVID-19 pandemic. The results presented show that epidemic risk perception is positively correlated with perceived stress, depression and anxiety.

There is a delay effect between the contact date of a carrier of the virus and the time when the contact is infectious or starts showing symptoms and also delay in testing and reporting during quarantine or isolation. This delay can be modeled in a stochastic [37] or a deterministic approach [38]. Studies have shown that many carriers of the virus do not show symptoms (they are called asymptomatic) but they are infected. This observation and that of many other factors have led researchers to model this kind of phenomenon using different approaches but adding a term that takes into consideration delay. [39] proposed a delay model to better understand the wide spread of the disease in Italy because at the beginning of the global pandemic there was a collapse of their health infrastructure and many deaths were recorded. The result presented by authors in this article present an insight for this delay and how it enhances the spread of the disease in Italy.

The COVID-19 pandemic has spread all over the world and it is important to study the propagation of information related to the spread of the disease, that is why in [40], the author presented a SIR model characterized by multi-layer networks which is a framework that incorporated the study of the interaction and awareness during epidemic spreading specifi-

cally. A related study proposed this type of framework in the case of COVID-19 outbreak in terms of misinformation of the people in relation to the virus, spread of rumors and fake news, and also the role of media (social and print) in the epidemic spread [41, 42, 43, 44]. They proposed that all these factors can affect the spread of the disease and the population opinion based on the information it accesses.
As the virus spread increases, government, health policy makers and experts suggested different prevention strategies which have helped to mitigate the spread of the disease and reduce the basic reproduction rate per time. Models incorporating these strategies have been developed, and in [45, 46] authors formulated a SIR model which was based on social distancing, a measure introduced by government in many countries to control COVID-19 disease spread. Recently, vaccines to effectively tackle the pandemic were produced after many trials by researchers. Vaccines such as Moderna, Pfizer, AstraZeneca, Novavax, Johnson and Johnson are many authorized in Europe which have been administered to people based on age and the number of doses depends on the efficacy of the vaccine because research proves that immunity decreases after some time reason why the third dose was introduced to better tackle the virus. Mathematical and statistical modeling of the pandemic taking into account vaccination has been studied. Author in [47] predicts the effectiveness of the vaccination, [48] used Geo-spatial model with Gillespie algorithm to investigate COVID-19 spread and vaccination, [49] analyzed and developed SEIR model for COVID-19 pandemic with vaccination and non singular kernel, [50] investigates the effectiveness of vaccination in the United States using predictive epidemic model. Authors were able to demonstrate that the implementation of vaccination depends on other social interventions introduced to reduce transmission rate at the population level.
There are other models that have been used to study COVID-19 disease. They are SEIR model with quarantine and fatality compartment [51], lattice model [52], a recursive model [53], computational model [54], networks based model [55, 56], tailored model [57], growth rate model [58], parameter estimation model [59], Weibull distribution model [60], Markov-chain model [61], in-homogeneous spatial model [62], fuzzy model [63], hybrid intelligent model [64], case based rate reasoning model [65], forecasting model [66], complex mathematical model [67], time series model [68], and hierarchical epidemic risk model [69]. All these models were used to predict, forecast and understand COVID-19 spread dynamics.
To conclude this section, it has been shown in this section that a lot of models have been developed and applied to predict and forecast the pandemic at different waves, by taking into account different control measures and interventions (quarantine, use of face-mask, social distancing and vaccination) introduced to avoid escalation of the disease. The estimation of the basic reproduction number, demographic models and socio-economic impacts of the pandemic, statistical and mathematical approaches were discussed in some articles. The goal of this thesis is to contribute significantly to the modeling of COVID-19 pandemic by developing, applying and suggesting new models using different approaches.

1.3 Thesis Outline

The first chapter is based on the general introduction. Chapter 2 deals with the state of art and the methodologies used in the thesis. Chapter 3 and Chapter 4 deals with the results obtained from the mathematical and statistical modeling of the COVID-19 outbreak which are from the scientific publications (published and in preparation) produced from the thesis. In the last chapter, the perspectives and conclusions derived from the thesis are discussed

and also the future works. Links to source code repositories used for the thesis can be found online at https://github.com/KayodeOshinubi. The thesis has been organized as follows:
Chapter 1 discusses the general introduction, systematic literature review, the thesis outline, and the list of scientific production from the thesis.
Chapter 2 discusses the mathematical methodology and the statistical methodology used throughout the work. The mathematical models developed in this thesis are an extension and improvement of the mathematical methodology presented. Also, the statistical methodology used throughout the work helps in predicting and forecasting the COVID-19 evolution at different waves.
Chapter 3 presents the mathematical models developed for the modeling of the COVID-19 outbreak dynamics and its several applications using a discrete and continuous approach and describing the importance of the models.
Chapter 4 deals with large scale analysis and the application of several statistical models to COVID-19 data-sets, time series and some socio-economic and demographic data because it is essential to understand the relationship between all these data and to properly describe the evolution of the pandemic and factors influencing the spread of the virus at different stages along the epidemic curve. Experimental results and analysis are presented
Chapter 5 is based on the conclusions and recommendations drawn from the thesis, description of the scientific contributions, and some future work.

1.4 Publications

The scientific publications, articles not yet published from this thesis, and some collaborative works that are not included in the thesis are listed in this section. Also, a part of the publications listed are conferences proceedings and posters presented. These publications also include collaborations with researchers across different countries. Publications from this research work cut across journals in applied mathematics, statistics, computational science and infectious disease/epidemiology. The publications are listed as follows:

Peer reviewed papers

- J. Demongeot, K. Oshinubi, M. Rachdi, H. Seligmann, F. Thuderoz and J. Waku. Estimation of Daily Reproduction Rates in COVID-19 Outbreak. *Computation*, 9(10):109, 2021. doi: 10.3390/computation9100109.

- K. Oshinubi, M. Rachdi and J. Demongeot. Analysis of daily reproduction rates of COVID-19 using Current Health Expenditure as Gross Domestic Product percentage (CHE/GDP) across countries. *Healthcare*, 9(10):1247, 2021.
doi: 10.3390/healthcare9101247.

- K. Oshinubi, F. Al-Awadhi, M. Rachdi and J. Demongeot. Data Analysis and Forecasting of COVID-19 Pandemic in Kuwait Based on Daily Observation and Basic Reproduction Number Dynamics. *Kuwait J. Sci.*, Special Issue:1-30, 2021.
doi: 10.48129/kjs.splcov.14501.

- K. Oshinubi, M. Rachdi and J. Demongeot. Modeling of COVID-19 pandemic vis-à-vis some socio-economic factors. *Front. Appl. Math. Stat.*, 7:786983, 2022.
doi: 10.3389/fams.2021.786983.

- K. Oshinubi, M. Rachdi A. C. Amakor, J.P. Olumuyiwa, A.A. Ayoade and J. Demongeot. Approach to COVID-19 Time Series Data Using Deep Learning and Spectral Analysis Methods. *AIMS Bioengineering*, 9(1):1-21, 2021. doi: 10.3934/bioeng.2022001.

- K. Oshinubi, F. Ibrahim, M. Rachdi and J. Demongeot. Functional Data Analysis: Application to Daily Observation of COVID-19 Prevalence in France. *AIMS Mathematics*, 7(4):5347-5385, 2022. doi: 10.3934/math.2022298.

- K. Oshinubi, S. BuHamra, N. Alkandari, J. Waku, M. Rachdi and J. Demongeot. Age Dependent Epidemic Modelling of COVID-19 Outbreak in Kuwait, France and Cameroon. *Healthcare*, 10(3):482, 2022. doi: 10.3390/healthcare10030482.

- J. Waku, K. Oshinubi and J. Demongeot. Maximal reproduction number estimation and identification of transmission rate from the first inflection point of new infectious cases waves: COVID-19 outbreak example. *Mathematics and Computers in Simulation*, 198:47-64, 2022. doi: 10.1016/j.matcom.2022.02.023.

- K. Oshinubi, C. Fougère and J. Demongeot. A model for the lifespan loss due to a viral disease: example of the COVID-19 outbreak. *Infect. Dis. Rep.*, 14(3):321-340, 2022. doi: 10.3390/idr14030038.

Published collaborative works

- J. Demongeot, K. Oshinubi, M. Rachdi, L. Hobbad, M. Alahiane, S. Iggui, J. Gaudart and I. Ouassou. The application of ARIMA model to analyse COVID-19 incidence pattern in several countries. *J. Math. Comput. Sci.*, 12:10, 2022. doi: 10.28919/jmcs/6541.

- A. Glaría, R. Salas, S. Chabert, P. Roncagliolo, A. Arriola, G. Tapia, J. Plaza, M. Salinas, H. Zepeda, K. Oshinubi and J. Demongeot. A stepforward to formalize Tailored to the Problem Specificity Mathematical Transforms: Use of ELM and explorations with new Biosignals. *Front. Appl. Math. Stat*, 8:855862, 2022. doi: 10.3389/fams.2022.855862.

- M. Jelassi, K. Oshinubi, M. Rachdi and J. Demongeot. Epidemic dynamics on social interaction networks. *AIMS Bioengineering*, 9(4): 348-361, 2022. doi: 10.3934/bioeng.2022025.

Conference proceedings, contributed talk and poster

- J. Demongeot, K. Oshinubi, M. Rachdi and H. Seligmann. Geoclimatic, Demographic and Socioeconomic determinants of the COVID-19 prevalence. *EGU General Assembly Conference Abstracts*, EGU21-7976, 2021. doi: 10.5194/egusphere-egu21-7976.

- K. Oshinubi, J. Demongeot and M. Rachdi. The biphasic pattern of the virulence curve of the first and second wave of COVID-19. *Coronavirus research conference proceedings*, 2021.

- K. Oshinubi, M. Rachdi and J. Demongeot. Age dependent modeling and its application to COVID-19 outbreak. *BAMC conference proceedings*, 2022.

- K. Oshinubi, M. Rachdi and J. Demongeot. Mathematical and Statistical Modeling of COVID-19 Outbreak. *Journée des doctorants de l'EDISCE*, 2022.

Chapter 2

State of the Art

2.1 Mathematical Methodology

In this section, we present the mathematical methodologies used in the thesis for the modeling of the COVID-19 disease. These methodologies include Bernoulli's epidemiological model, D'Alembert's epidemiological model, Ross and McKendrick SIR model, Leslie Model and Usher's Model.

2.1.1 Bernoulli's Epidemiological Model

Daniel Bernoulli [1] developed a mathematical model by taking into consideration the dynamics of smallpox infection. In the model, the whole population was divided into susceptibles and immunized where the probability of a newborn to be susceptible is denoted by p and the probability of being immunized to the smallpox infection is denoted as q.

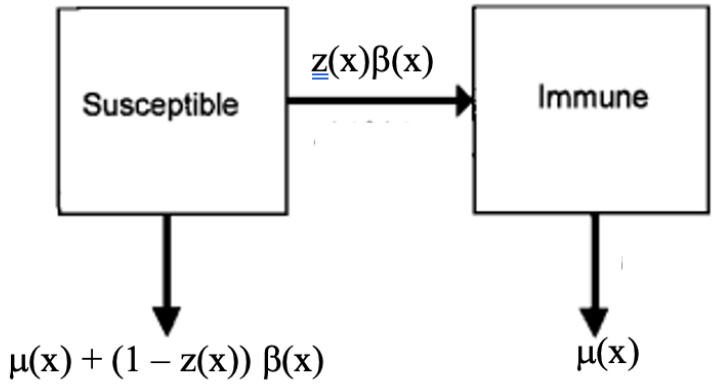


Figure 2.1: Structure of Bernoulli's epidemiological model for a population divided into susceptible and immunized. At age x, the probability of surviving the infection is $z(x)$, transmission rate is $\beta(x)$ and natural death rate is $\mu(x)$.

At age x, the natural death rate is denoted by $\mu(x)$, and the transmission rate is denoted by $\beta(x)$. Only a fraction of susceptibles $z(x)$ survives to become immune while the rest $b(x) = 1 - z(x)$ dies due to the infection which means that $b(x)$ is the fatality rate. We will refer to it as case fatality because it is not a rate (with dimension per unit of time) but a probability, i.e. a dimensionless quantity. Let $p(x)$ denote the likelihood of a newborn being

alive and susceptible at age x. Figure 2.1 shows the structure of Bernoulli's epidemiological model. The differential equation is then satisfied by $p(x)$:

$$\frac{dp}{dx} = -(\beta(x) + \mu(x))p,$$

with the initial condition $p(0) = 1$.
The probability $q(x)$ to be immunized for life and be alive is represented by the differential equation below:

$$\frac{dq}{dx} = (1 - b(x))\beta(x)p(x) - \mu(x)q,$$

with the initial condition $q(0) = 0$.
Daniel Bernoulli used the logistic model as a phenomenological approach to fit data to the model he developed above corresponding to the daily new cases of the smallpox infection. He described the disease's endemic state following a wave, but he did not propose a model for describing the endemic dynamic that is now considered to be a stationary stochastic process with constant average and variance, nor did he use a precise method to detect critical boundary times corresponding to ruptures between epidemic and endemic phases.

2.1.2 D'Alembert's Epidemiological Model

D'Alembert [70] developed in 1761 an alternative method for dealing with competing risks of death, which is also applicable to non-infectious diseases. D'Alembert developed an alternative solution in approaching epidemiological modeling developed by Daniel Bernoulli which nowadays one would call a non-parametric approach in contrast to the Bernoulli parametric model [71]. Figure 2.2 shows the structure of D'Alembert's epidemiological model. D'Alembert approach is quite general and is not restricted to an immunizing infectious disease. Suppose $\lambda(x)$ denote the death rate at age x due to a particular infectious disease and the natural death rate at age x is denoted as $\mu(x)$. Let $\psi(x)$ denote the rate at which deaths due to a specific cause are recorded for people who die at age x. Then we have:

$$\psi(x) = \lambda(x)l(x),$$

where the survival function at age $x, l(x)$, is defined by:

$$\log l(x) = \sum_{y \leq x} \log(1 - \mu(y)).$$

If the survival function $l(x)$ is known, the mortality due to the disease can be calculated by dividing $\psi(x)$ by $l(x)$. The d'Alembert's epidemiological model is unquestionably more general than the Bernoulli epidemiological model, which is limited to immunizing against infections. The Bernoulli's epidemiological model, on the other hand, provides significantly more insight for the interpretation of infectious disease data.

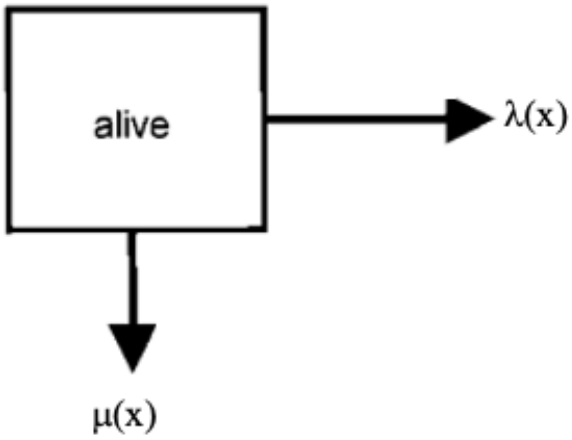


Figure 2.2: Structure of D'Alembert's epidemiological model where the two deaths rates - $\mu(x)$ (natural death rate) and $\lambda(x)$ (death rate due to the infectious disease) are assumed to act independently.

2.1.3 Ross and McKendrick SIR Model

The most common SIR model which is an improvement of Bernoulli epidemiological model is the SIR Model by Ross and then by McKendrick [72, 73], which was assisted by W.O. Kermack. The Kermack-McKendrick model is a SIR model for the number of people infected by an infectious disease in a closed population over time. It was proposed to explain the rapid rise and fall in the number of infected cases observed in disease outbreaks. This model represents diseases against which people have a lifetime immunity and it does not consider the possibility of re-infection with the disease which is also an important factor that should be considered. The model is expressed as:

$$\begin{aligned} \frac{dS}{dt} &= \delta S + \delta I + (\delta + \gamma)R - \beta SI - \mu S, \\ \frac{dI}{dt} &= \beta SI - (\mu + \nu)I, \ \frac{dR}{dt} = \nu I - (\mu + \gamma)R \end{aligned}$$

where S is the size of the susceptible, I the size of the infectious, and R the size of the recovered, with the total population N defined as $N = S + I + R$. β is the transmission rate, δ the birth rate (supposed to be equal to the death rate μ), γ is the loss of immune resistance, and ν is the immunization rate. Determining the epidemiological threshold is a critical component in determining whether or not to implement lockdown. This decision would also be based on the disease's fatality, which is beyond the scope of the SIR models. The basic reproduction number which is the epidemiological threshold is expressed as:

$$R_0 = \frac{\beta N}{\nu + \delta}$$

The model can also be used for age dependent modeling by defining age classes between 1 and m and by denoting $S_j, I_j, R_j, j = 1, 2, \ldots, m$, each subpopulation of S, I, and R at age j, we can define at any stationary state (S^*, I^*, R^*) the probabilities for a newborn individual of being alive and either susceptible, infectious or immune at age j by the following formula:

$$u^*(j) = \frac{S_j^*}{P^*}, v^*(j) = \frac{I_j^*}{P^*}, w^*(j) = \frac{R_j^*}{P^*}, \text{ where } P* = \sum_{i=1,\ldots,m} (S_i + I_i + R_i)$$

These formula make the link between the Bernoulli and the Ross and Kermack-McKendrick epidemiological models.

2.1.4 Leslie Model

Leslie model which is sometimes called Leslie matrix is the first discrete population growth model used for age structured modeling and it is well known for projecting how age is distributed in a particular population. The model was proposed by Leslie in 1945 [74, 75]. It majorly considers one sex (female), fecundity and deaths among the sub-population (different age classes) with a population of closed migration. Leslie model uses the age pyramid vector $p(t) = (p_j(t))_{j=1,2,\ldots,m}$, where $p_j(t)$ represents the size of the age class j at time t, with j ranging from the birth age 1 to the maximal death age m, whose discrete dynamics is governed by the matrix equation given as:

$$p\,(t) = Lp\,(t-1) \text{ with } L = (l_{jk}) = \begin{bmatrix} f_1 & f_2 & f_3 & \ldots & \ldots & f_m \\ b_1 & 0 & 0 & \ldots & \ldots & 0 \\ 0 & b_2 & 0 & \ldots & \ldots & 0 \\ \vdots & \vdots & \ddots & \ldots & \ldots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ldots & \vdots \\ 0 & 0 & 0 & \ldots & b_{m-1} & 0 \end{bmatrix}$$

where $\forall\ j = 1, \ldots, m, b_j = 1 - \mu_j \leq 1$ (where μ_j is the death rate at age j) is the survival probability between ages j and $j+1$ and f_1 is the birth rate at age j (i.e., the mean number of offsprings from an individual at age j). The dynamic stability of the Leslie system for a distance d is quantified by the tendency to return to its stationary pyramid $p*$ after a perturbation $p^* + q$, such as $d(p^*, L^m(p^* + q)) < Ke^{-mD}$ tends exponentially to 0, when m goes to infinity, the parameter D being the stability module. For the L^2 distance to the stationary pyramid $p^*, D = |\lambda - \lambda'|$, absolute value of the difference between the dominant and sub-dominant eigenvalues of L, i.e., λ and λ' ($\lambda = e^r$, where r is the Malthusian growth rate, and p^* is the left eigenvector of L corresponding to λ). For the distance (known as symmetrized divergence) of Kullback–Leibler, $D = kH$, where H is the entropy of p^* and k is a constant [76, 77].

The Leslie model has the advantage of not requiring a stable age distribution for valid population projections. It also allows for sensitivity analysis to see how changing certain age classes vital rates affects population size and age structure. It can also incorporate density-dependence, i.e., dampen values in the matrix to account for density-dependent factors limiting population growth, and it can derive useful mathematical properties from matrix formulas, such as stable-age distribution and finite rate of population change (i.e., the growth parameter λ). The model's disadvantages are that it requires a large amount of data (i.e., age-specific data on survival, fecundity, and population structure) and estimating fecundity is difficult atimes.

2.1.5 Usher's Model

An improvement of the Leslie matrix is the Usher's model which is sometimes refers to as Usher's matrix [78] which is widely used to predict population dynamics and to predict the dynamics of size-structured or age-structured populations. The model is formulated in

which it is possible to remain in the same biological age corresponding to an increase of the longevity (a rejuvenation) or to pass over a biological age state corresponding to an acceleration of ageing between $t-1$ and t as modeled by Usher using the vector of age class sizes $u(t) = [u_1(t), \cdots, u_m(t)]$, whose discrete dynamics is ruled by matrix equation $u(t) = Uu(t-1)$, where:

$$U = (u_{ij}) = \begin{bmatrix} f_1 + v_1 & f_2 & f_3 & \cdots & f_{m-1} & f_m \\ \alpha_1 & v_2 & 0 & \cdots & \cdots & 0 \\ 0 & \alpha_2 & v_3 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & . & . & \vdots \\ \vdots & \vdots & \vdots & \ddots & v_{m-1} & \vdots \\ 0 & 0 & 0 & \cdots & \alpha_{m-1} & v_m \end{bmatrix}$$

and where $f_1, f_2, \ldots, f_m$ are the fertility rates in age classes, v_i is the probability to remain in state i, α_i is the probability to go from state i to state $i+1$, $v_i+\alpha_i = 1-\mu_i \leq 1$, $\forall\, i = 1, \ldots, m$, where μ_i is the death rate at age i. and v_j. The probabilities P_i belong to the interval [0,1]: $0 \leq P_i \leq 1 \ \forall \ i = 1, \ldots, m$. We can also say that $(v_i + \alpha_i)$ is the general probability of survival. The dynamical stability modulus for the L^2 distance between the current age pyramid $[u_i(t)/\sum_{j=1,\ldots,m} u_j(t)]$ and the stationary age pyramid w is given by $e^{-|\lambda-\lambda'|}$, where λ(resp. λ') is the dominant (resp. sub-dominant) eigenvalue of the matrix U. In general, the explicit calculation of the eigenvalues of the Usher matrix is not possible.
Many theoretical properties of the Usher model are known, particularly its asymptotic behavior: the evolution to the stationary state is exponential and is characterized by the growth rate and the stationary distribution [79, 80]. Figure 2.3 shows the life cycle representation of the Usher projection matrix model.

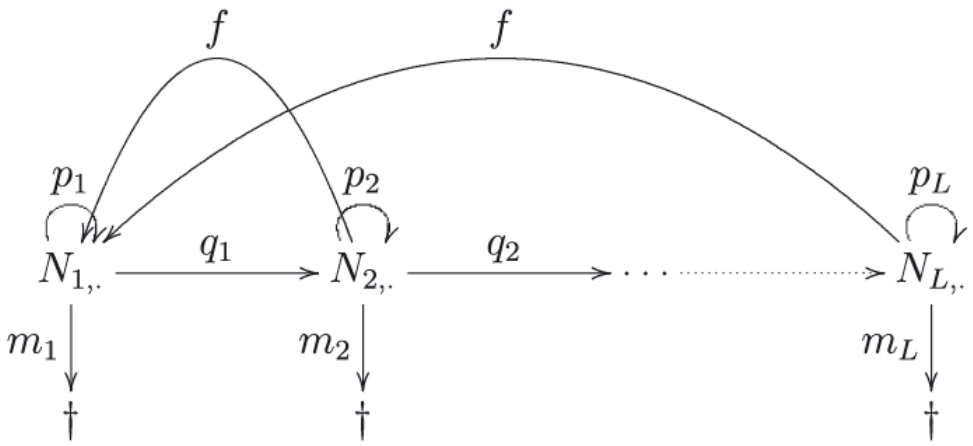


Figure 2.3: Life cycle representation of the Usher projection matrix model, where p_i is the probability for an individual to stay in class i, q_i is the probability to move up from class i to $i+1$, m_i is the probability of dying and f is the average fertility rate supposed to be constant (after [79]).

2.2 Statistical Methodology

This section deals with the statistical methodologies used in the thesis for the modeling of the COVID-19 disease. These methodologies include time series, machine learning (supervised and unsupervised), deep structured learning (feed-forward neural network and recurrent neural networks), spectral analysis and functional data analysis. In Figure 2.4, we present the block diagram for the statistical methodology applications study.

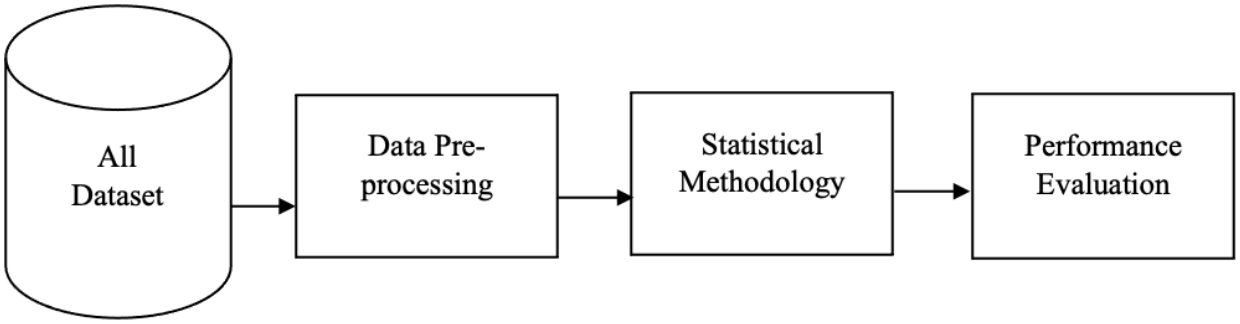


Figure 2.4: Block diagram of the statistical methodology.

The data used throughout this study were collected from public database for all epidemiological, demographic, and socio-economic data and references were provided as appropriate in various sections and subsections.

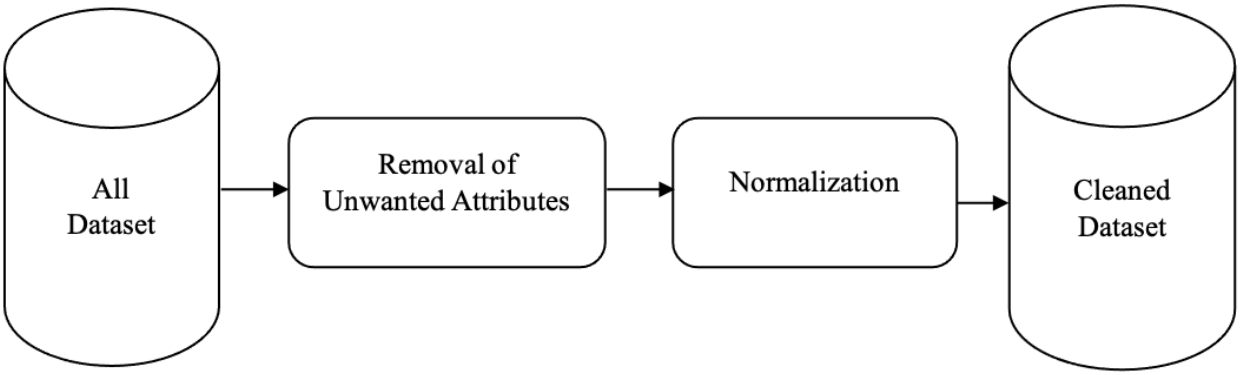


Figure 2.5: Data pre-processing of all data-set used in the study.

The data-set used raw file was overflowed by a large quantity of information collected. The raw file processing ensured the right attributes to be mined for this study. The data-set was pre-processed and cleaned on Microsoft Excel while some of the data-set was converted to Comma Separated Values (CSV) files. The data cleaning processes ensure the removal of unwanted attributes and the quality data-set creation for an excellent analysis. The data were normalized to generate a cleaned data-set. Figure 2.5 presents the steps involved in data preparation and cleaning before applying the statistical techniques. In subsequent section, we present the mathematical formulation of all the statistical methodologies used in this thesis having in mind that most of these models are already inbuilt in well known software like Python® and R which were employed for the experimental work.

2.2.1 Time Series Modeling

Most of the data-sets used in this research are time series data and we shall give details about some of the time series modeling used to predict and forecast our data-sets. Time series data set can be hourly, daily, monthly or yearly because one of the variables has to do with time and it consists of four components which are trend, cyclical, seasonal and irregular and it depends on the component of interest you want to investigate in the modeling. It can be represented as follows:

$$\{p_t\}_{t=1,\ldots,T}$$

where p_t is a random variable which represents a repeated observations.

Auto-regressive integrated moving average

The auto-regressive integrated moving average model, abbreviated as ARIMA, is a well-known and widely used statistical technique for time series forecasting that uses the form of the general equation presented in Equation (2.1) to forecast future values of a time series based on previous values. ARIMA can model non-seasonal time series with non-white noise patterns. Box and Jenkins introduced the ARIMA model in 1970 [81]. ARIMA model has proven to be capable of producing accurate short-term forecasts. This model is based on the idea that a variable's future value of a time series is determined by its past values and errors. The model equation is as follows:

$$p(t) = \sum_{i=1,\ldots,r} a(t)p(t-i) + \varepsilon(t), \tag{2.1}$$

where $p(t)$ is the time series real value at a time t and ε is a random residue also known as noise, whose variance is to minimize. The auto-correlation curve is obtained by calculating the correlation $A(k)$ between $p(t)$ and the $p(t-k)$'s (t belonging to a moving time window) by using the formula:

$$A(k) = \frac{E[p(t)p(t-k)] - E(p(t))E(p(t-k))}{\sigma(p(t)\sigma(p(t-k))}. \tag{2.2}$$

where E denotes the empirical expectation calculated on a chosen moving time window of length m and σ is the standard deviation.The auto-correlation function A allows examining the serial dependence of the $p(t)$'s.
The noise (white if it is Gaussian) ε is a stochastic process and must obey the following rules:

$$E(\varepsilon(t)) = 0$$

$$V(\varepsilon(t)) = \delta_\varepsilon^2 < +\infty$$

$$Cov(\varepsilon(r), \varepsilon(r-k)) = 0, \forall\ r \text{ and } k \neq 0$$

where V denotes the variance and Cov the auto-covariance of the stochastic process ε.
In this thesis, we used the classical ARIMA (r, d, m) model, where r means the order of auto-regression, d the degree of trend difference (related to the polynomial character of the trend and discrete analog to the order of a differential equation) and m the order of moving average.

Exponential smoothing modeling

There are fifteen types of exponential smoothing modeling as stated in [82] which depend strictly on underlying error (Multiplicative - M or Additive - A), type of trend (None - N, Additive - A or Damping - D) and type of seasonality (None - N, Additive - A or Multiplicative - M). The simple exponential smoothing model which has no trend nor seasonality but a slowly varying mean for a time series p_t is given as follows:

$$p_{t+h} = l_t \tag{2.3}$$

$$l_t = \alpha p_t + (1-\alpha) l_{t-1}$$

where p_{t+h} the h-step ahead forecast, l_t is the level of the time series and α is the parameter which determines the degree of the smoothing and is always estimated by minimizing the sum of the squared prediction errors.

It is important to first describe the underlying error models which can be described for simple exponential smoothing model with an additive error as:

$$p_{t+h} = l_{t-1} + \varepsilon_t \tag{2.4}$$

$$l_t = l_{t-1} + \alpha \varepsilon_t$$

and for the simple exponential smoothing model with a multiplicative error, is given as:

$$p_{t+h} = l_{t-1}(1+\varepsilon_t) \tag{2.5}$$

$$l_t = l_{t-1}(1+\alpha\varepsilon_t)$$

The improvement of Equation (2.3) with a trend behavior of the time series which can give rise to a model like MAN (exponential smoothing with additive trend without seasonal component and with a multiplicative underlying model) is expressed as:

$$p_{t+h} = l_t + \sum_{i=1}^{m} \Phi^i a_t$$

$$\begin{aligned} l_t &= \alpha p_t + (1-\alpha)(l_{t-1} + \Phi a_{t-1}) \\ a_t &= \beta(l_t - l_{t-1}) + (1-\beta)\Phi a_{t-1} \end{aligned} \tag{2.6}$$

where β is the parameter that determines the rate of change in trend and Φ is the parameter that takes into consideration the damping of the trend. Other parameters are defined as previously. If we set $\Phi = 0$, we will have a simple exponential model without trend (e.g. ANN or MNN) but if $\Phi = 1$, it gives an exponential model with trend (e.g. AAN).
Another improvement of Equation 2.3 is to add seasonality effects to the model. For additive seasonality we have:

$$p_{t+h} = l_t + \sum_{i=1}^{m} \Phi^i a_t + s_{t-n+h_n}$$

$$\begin{aligned} l_t &= \alpha(p_t - s_{t-n}) + (1-\alpha)(l_{t-1} + \Phi a_{t-1}) \\ a_t &= \beta(l_t - l_{t-1}) + (1-\beta)\Phi a_{t-1} \\ s_t &= \gamma(p_t - l_{t-1} - \Phi a_{t-1}) + (1-\gamma)s_{t-n} \end{aligned} \tag{2.7}$$

for multiplicative we have:

$$
\begin{aligned}
p_{t+h} &= (l_t + \sum_{i=1}^{m} \Phi^i a_t) s_{t-n+h_n} \\
l_t &= \alpha(p_t/s_{t-n}) + (1-\alpha)(l_{t-1} + \Phi a_{t-1}) \\
a_t &= \beta(l_t - l_{t-1}) + (1-\beta)\Phi a_{t-1} \\
s_t &= \gamma(p_t/(l_{t-1} - \Phi a_{t-1})) + (1-\gamma)s_{t-n}
\end{aligned}
\tag{2.8}
$$

. where $h_n = \lfloor (h-1) \mod n \rfloor + 1$. The seasonality parameter is γ and the number of season is given as n which for instance for monthly data-set verifies: $n = 12$. Other parameters are defined in the same way as before.
Table 2.1 gives a summary of the symbols we have described earlier for all the exponential smoothing models.

Table 2.1: Exponential smoothing models.

Trend	**Seasonality**		
	None(N)	**Additive(A)**	**Multiplicative(M)**
None(N)	ANN/MNN	ANA/MNA	MNM
Additive(A)	AAN/MAN	AAA/MAA	MAM
Damping(D)	ADN/MDN	ADA/MDA	MDM

Prophet forecasting modeling

The Prophet forecasting modeling is an improved version of ARIMA model and exponential smoothing model. It uses three main model components which are trend, seasonality and holidays. It is described by the model equation below:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon(t), \tag{2.9}$$

where $g(t)$ is the trend, a piece-wise linear or logistic growth curve for modeling the non-periodic changes in time series defined for the logistic case by:

$$g(t) = C/(1 + \exp(-k(t-z))),$$

where C is carrying capacity, z is the growth rate and m is the offset parameter.

Similarly, for the piece-wise linear case, we have:

$$g(t) = (s_0 + s_1 + \cdots + s_j) \times t + (C_0 + C_1 + \cdots + C_j),$$

where j is the last change point before t.
Other parameters in Equation (2.9) are as follows: $s(t)$ represents the seasonality, $h(t)$ is the effect of holidays with irregular schedules and $\varepsilon(t)$ is the error term which accounts for any unusual changes or noise not accommodated by the model [83].
Prophet model forecasts seasonality effects using the Fourier series, and the seasonality models are specified as periodic functions of t. The arbitrary smoothing of seasonal effects with a scaling time variable using Fourier series is of the form:

$$s(t) = \frac{a_0}{2} + \sum_{m=1}^{\infty} a_n \cos\frac{2m\pi t}{P} + b_n \sin\frac{2m\pi t}{P},$$

where P is the period for fitting the seasonality and the parameters a_m and b_m can be estimated from the series. The seasonality is multipliable by the trend if the argument "multiplicative" is specified then we have:

$$s^*(t) = \begin{cases} g(t) \times s(t) \text{ if multiplicative} \\ s(t), \text{ otherwise.} \end{cases}$$

If we take the period $P = 1$ and 9, by considering additive seasonality (as opposed to multiplicative), we have:

$$s(t) = s_1^*(t) + s_9^*(t) = s_1(t) + s_9(t) = \sum_{m=1}^{9} a_n \cos \frac{2m\pi t}{1} + b_n \sin \frac{2m\pi t}{1} + \sum_{m=1}^{5} a_n \cos \frac{2m\pi t}{9} + b_n \sin \frac{2m\pi t}{9}$$

Also, $h(t)$ can be a multiplicative seasonality and if we introduce a trainable coefficient σ_m then it can be defined as follows:

$$h(t) = \sum h_m^*(t)$$

where,

$$h_m(t) = \sigma_m h(t)$$

$$h_m^*(t) = \begin{cases} g(t) \times h_m(t) \text{ if multiplicative} \\ h_m(t), \text{ otherwise.} \end{cases}$$

for each m.

Spectral Analysis

The decomposition of a time series into underlying sine and cosine functions of different frequencies allowing us to identify those frequencies that appear to be strong or very important is called spectral analysis. Another way to think about spectral analysis is as a linear multiple regression problem, with the dependent variable being the observed time series and the independent variables being the sine and cosine functions of all possible frequencies. Spectral analysis always identifies the correlation of sine and cosine functions of different frequencies with the observed data, and if a large correlation is found, one can conclude that the respective function has a strong periodicity in the real data. Spectral analysis is useful for analyzing stationary time series and identifying noise-corrupted periodic signals [84].

The periodogram is one of the most important tools in spectral analysis because it is an estimation called spectral density. It quantifies the contributions of individual frequencies to time series regression and is denoted as $P_k = a_k^2 + b_k^2$, where P_k is the periodogram, a_k cosine parameter, and b_k sine parameter values at frequency k (for $k = 1, 2, \ldots, n/2$). The periodogram values can be interpreted in terms of the variances of the real data at the respective frequencies. A Fourier line spectrum is a plot of P_k (as spikes) versus k. The periodogram is created by joining the tips of spikes in the Fourier line spectrum to form a continuous plot and scaling it to an area equal to the variance. The periodogram can also be smoothed by using different Daniell kernel which can be log or sine or cosine.

Another important tool in spectral analysis is multitapering which is an extension of single taper approaches that involves dividing the data into overlapping subsets that are individually tapered and then Fourier transformed. The individual spectral coefficients of each subset are

averaged to reduce variance. The method of subdividing the data is defined by a set of tapers indexed by $m = 1 \cdots M$; the estimated spectral density at frequency f is expressed as follows:

$$Z(f) = \frac{\sum_{m=0}^{M-1} \lambda_m Z_m(f)}{\sum_{m=0}^{M-1} \lambda_m},$$

where $Z_m(f)$ is the estimated spectral density using taper m and λ_m are weights for each tapered spectral density estimate.
Also, harmonic regression analysis of the frequencies of the time series is important. Suppose we have a time series p_t which contains a periodic component, a phenomenon which spectral density is known for, the model will be of the form:

$$p_t = \mu + Zcos(ft + d) + \varepsilon_t,$$

which can further be expanded to be of the form:

$$p_t = \mu + a\cos(ft) + b\sin(ft) + \varepsilon_t,$$

where μ is the mean of the series, Z is the amplitude of variation and d is the horizontal offset. All other parameters remain defined as earlier stated.
This nonlinear model can be fitted as a linear regression model with two independent variables, $P = \cos(ft)$ and $Q = \sin(ft)$. The regression coefficients are $J = a$ and $K = b$. In practice, the variation in a time series can be modeled as the sum of several different individual sinusoidal terms occurring at different frequencies. The generalization of this model to the sum of k frequencies, to which we can optionally add a trend term to obtain the forecasting equation, can be expressed as follows:

$$p_t = \mu + zt + \sum_{m=1}^{k} a_m \cos(f_m t) + \sum_{m=1}^{k} b_m \sin(f_m t) + \varepsilon_t.$$

2.2.2 Supervised Machine Learning Modeling

Supervised machine learning modeling involves classification and regression supervised learning problems and processes. In most cases, you give specific instructions to what the machine will do. Figure 2.6 shows the flowchart of used for supervised machine learning modeling, analysis and the process involved while Figure 2.7 shows the flowchart for k-fold cross-validation strategy which is a re-sampling method to evaluate supervised machine learning models on a limited dataset sample.

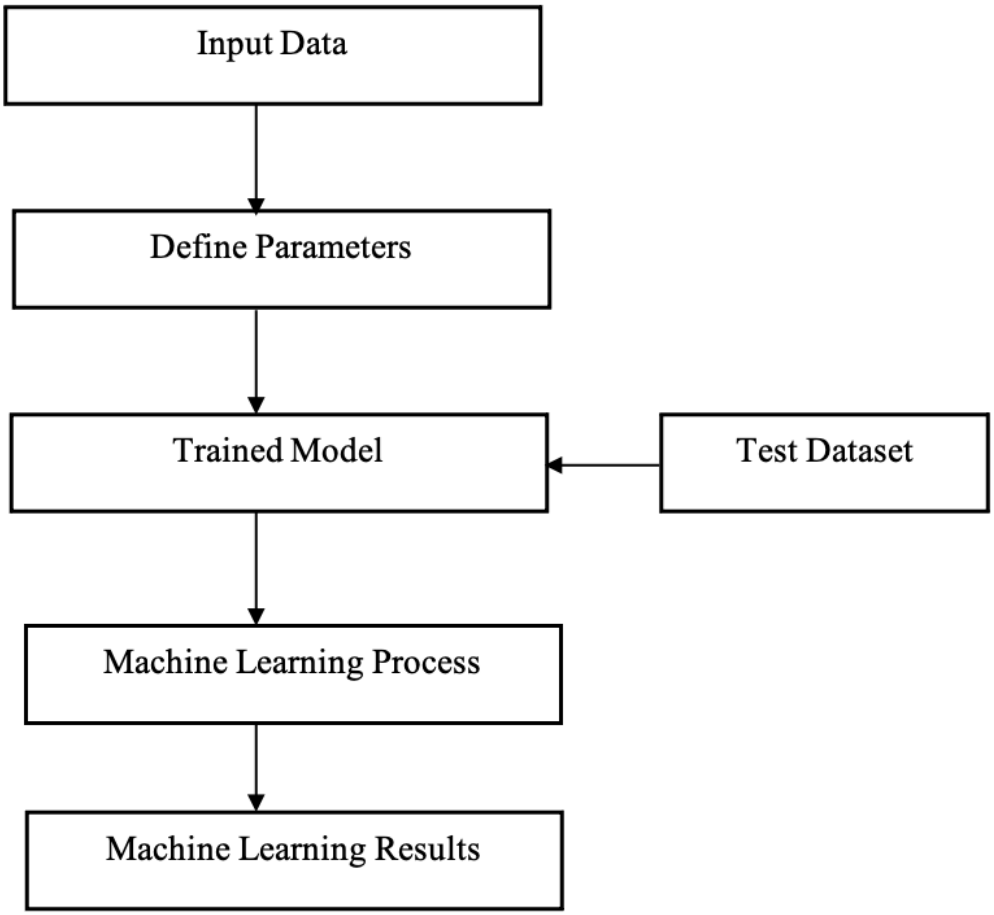


Figure 2.6: Flowchart of supervised machine learning technique.

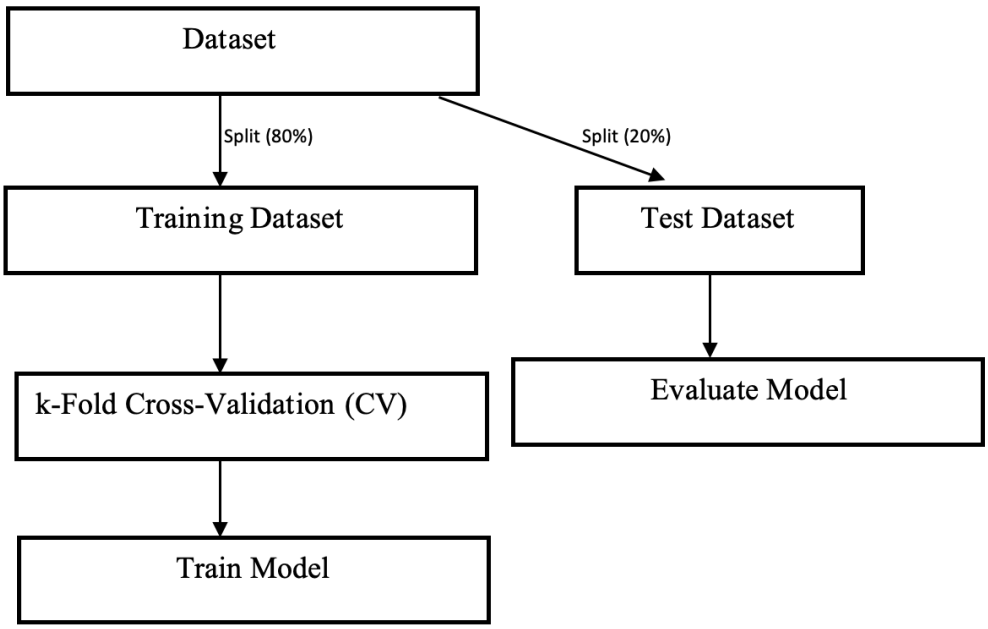


Figure 2.7: Flowchart for k-fold cross-validation strategy to evaluate supervised machine learning technique.

Regression modeling

Here we describe the linear regression model, log-linear regression model, generalized linear models, polynomial regression (order ≥ 2) model, regression splines, multivariate ordinary least square model, ridge regression model and lasso regression model.
Linear regression model uses some historic data of independent and dependent variables and consider a linear relationship between both. The model is of the form:

$$y = bx + a,$$

where a is the intercept, b is the slope, x is the independent variable and y is the dependent variable.
Log-linear regression model is coming from the common exponential model which is given as:

$$y = ae^{bx}, \tag{2.10}$$

where y is the daily number of incidence in the case of modeling COVID-19 outbreak which is the subject of the thesis, x is the number of days, b is the slope and a is a constant, and the log-linear model is derived by taking the *log* of both sides of Equation (2.10) which can be written as:

$$\log y = \log a + bx.$$

It means that the log of the daily number of incidence will be used as the dependent variable.
Generalized linear models are models that fit a linear response function y_i as linear combination of the predictors x_{pi} to a data that is likely to have errors ε_i that are normally distributed [85]. It can be written as follows:

$$y_i = b_0 + \sum_p b_p x_{pi} + \varepsilon_i, \tag{2.11}$$

where y_i can be any exponential family distribution i.e., Binomial, Bernoulli, Gamma, Normal, Poisson etc. and $\varepsilon_i \sim N(0, \sigma^2)$ which means that the mean is equal to 0 and constant variance is σ^2. The intercept is b_0 and b_p is commonly estimated using ordinary least squares model.
Polynomial regression models use a similar approach as linear regression model but the dependent variable is modeled as a degree n ($n \geq 2$) polynomial in x. We can describe a cubic polynomial ($n = 3$) regression as follows:

$$y = b_1 x + b_2 x^2 + b_3 x^3 + b_0,$$

where b_1, b_2, b_3 are the weights of the coefficients of x. Other parameters are defined in the same way as before. To model other type of polynomial regression, the power of x will just be increased.
A spline is a smooth piecewise polynomial function with a fixed degree z defined and constructed by joining polynomials smoothly at values $\tau_i = 1, \ldots, m$. The values τ_i are known as knots or breakpoints, and they are given as follows:

$$t_1 = \tau_1 < \cdots < \tau_m = t_n.$$

Regression splines of order m with knots $\varepsilon_k, k = 1, \ldots, K$ is a piecewise-polynomial of order $m - 1$, which has a continuous derivatives up to order $m - 2$ whose basis functions is as

follows:

$$y = \sum_{n=1}^{K+m} \beta_n h_n(x) + \varepsilon.$$

The general formula for the truncated-power basis set is

$$h_k(x) = x^{k-1}, k = 1, \ldots, m$$

$$h_{k+m}(x) = (x - \varepsilon_k)_+^{m-1}, k = 1, \ldots, K$$

The regression spline can be linear or cubic but it is claimed that cubic splines are the lowest-order splines in which the knot-discontinuity is not visible to the naked eye. The set of linear or cubic splines with fixed knots is a vector space. There is almost never a good reason to go beyond cubic-splines. In practice the most widely used orders are $m = 1, 2$ and 4.
Multivariate ordinary least squares model allows us to test much more complex relations between variables. It can be can be represented as follows:

$$y = \beta_1 x_1 + \beta_2 x_2 \ + \cdots + \varepsilon, \tag{2.12}$$

where β_1, β_2 , ... are coefficients or weights, ε is the residual noise, y is the dependent variable and x_1, x_2 , ... are the independent variables.
Ridge regression is a modeling approach that can be used to reduce model complexity and false positive which may result from the classical linear regression model. The cost function for ridge regression is given as follows:

$$\sum_{i=1}^{m} \left(y_i - \sum_{j=0}^{p} \beta_j \times x_{ij} \right)^2 + \alpha \sum_{j=0}^{p} \beta_j^2,$$

for all $c > 0$, $\sum_{j=0}^{p} \beta_j^2 < c$, while α is the penalty term that regularizes the coefficients. If the coefficients take large values, the optimization function is penalized. Ridge regression puts constraints on the coefficients β.
In data modeling, Lasso (Least absolute shrinkage and selection operator) regression model helps to regularise data models and to select the best feature for the model. The cost function for Lasso regression is obtained in the same way as that of Ridge regression with little modification which is given as:

$$\sum_{i=1}^{m} \left(y_i - \sum_{j=0}^{p} \beta_j \times x_{ij} \right)^2 + \alpha \sum_{j=0}^{p} |\beta_j|.$$

Support vector machine

Support vector machine is a supervised machine learning model that draws a hyperplane relating to the data points and creates a boundary of possible data points (high and low) for prediction. The goal of the model is to find the maximum margin, which means to minimize error. A tolerance margin (ϵ) is set to approximate the model and minimize error, individualizing the hyperplane that maximizes the margin while keeping in mind that some of the error is tolerated. The mathematical formulation of the support vector regression where b and w are the coefficients is given as:

$$y = f(x, w) =< w, x >= \sum_{j=1}^{m} w_j x_j + b, \tag{2.13}$$

with $w_j > 0$. The support points minimize prediction error where $\|w\|$ is the magnitude of the vector w :

$$\min_w \frac{1}{2}\|w\|^2, \tag{2.14}$$

with constraints

$$\mid y_j - w_j x_j \mid \leq \varepsilon. \tag{2.15}$$

The error term is instead handled in the constraints, where we set the absolute error less than or equal to a specified margin, called the maximum error (ε). Parameters used for support vector machine are shown in Table 2.2.

Table 2.2: Model parameters for support vector machine.

Hyperparameter	**Nature/Value**
Type	Eps-regression
Kernel	Radial
Cost	1
Gamma	1
Epsilon	0.1
p-value	$< 2.2e^{-16}$

Canonical correlation analysis

Canonical correlation is an aspect of multivariate statistical modeling that is used to simultaneously correlate several metric dependent variables and several metric independent variables measured on or observed with similar experimental units. It finds linear combinations within a data set with the goal of maximizing the correlation between these linear combinations [86]. Mathematically, it can be expressed as two groups of n-dimensional variables

$$X = [x_i] = [x_1, x_2, \ldots, x_p]$$

and

$$Y = [y_i] = [y_1, y_2, \ldots, y_q],$$

where

$$x_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix}, y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iq} \end{bmatrix}.$$

The purpose of canonical correlation analysis is to find coefficient vectors

$$a_1 = (a_{11}, a_{21}, \ldots, a_{p1})^T \text{ and } b_1 = (b_{11}, b_{21}, \ldots, b_{q1})^T,$$

in order to maximize the correlation $\beta = corr(Xa, Yb)$, while $U = Xa$ and $V = Yb$. Linear combinations of X and Y components respectively, constitute the first pair of canonical predictors. Then, the second pair of canonical response can be found in the same way subject to the constraint that they are uncorrelated with the first pair of variables. By repeating this procedure, the $r = \min\{p, q\}$ pairs of the canonical response can be found and we will finally

get the matrix $A = [a_1, a_2, \ldots, a_r]$ and the matrix $B = [b_1, b_2, \ldots, b_r]$ to transfer X and Y to canonical responses U and V following the below expression:

$$U_{n\times r} = X_{n\times p}A_{p\times r}, V_{n\times r} = Y_{n\times q}B_{q\times r}$$

If X and Y are both centered, we can concatenate them and calculate the covariance matrix given as:

$$C = Cov([XY]) = \frac{1}{n-1}[XY][XY]^T = \begin{bmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{bmatrix},$$

where C_{xx} and C_{yy} are within the set of covariance matrices and $C_{xy} = [C_{yx}]^T$ is between the set of covariance matrices. The first canonical response a and b maximize the equation:

$$\beta = \frac{a^T C_{xy} b}{\sqrt{a a^T C_{xx}}\sqrt{b b^T C_{yy}}}$$

The generalized formulae to calculate other canonical responses a_i and b_i $\forall$ $i \geq 2$ can be expressed as follows:

$$\beta_i = \frac{a_i^T C_{xy} b_i}{\sqrt{a_i a_i^T C_{xx}}\sqrt{b_i b_i^T C_{yy}}},$$

subject to the constraint:

$$a_j a_i^T C_{xx} = 0 \ \forall \ j < i, b_j b_i^T C_{yy} = 0 \ \forall \ j < i.$$

2.2.3 Unsupervised Machine Learning Modeling

Principal component analysis

Principal component analysis is an exploratory data analysis technique which uses the observations in a data set say p for each m individuals or quantities [87]. A p m-dimension vector x_i of observations on the ith variable and a linear combination and the variance of the data set matrix denoted as X with minimum variance is defined as follows:

$$\sum_{i=1}^{p} b_i x_i = X b_i,$$

$$var(X b_i) = b^T S b_i;$$

where b is a vector whose components are constants which are commonly called eigenvectors, S is the sample covariance matrix associated with the data set. The linear combinations Xb_i are called principal components of the data set. The elements of the eigenvectors b_i are called principal components loadings and elements of the linear combinations Xb_i are called the principal components scores which are values each m of the individuals or quantities will score on principal components. The first principal component gives the most information in principal component analysis.

Clustering analysis

Clustering is a data analysis technique that divides objects into different clusters by measuring distances and identifying individual clusters [88]. K-means and Hierarchical techniques are two main types of clustering analysis that will be discussed in this section. Figure 2.8 shows the flowchart of clustering techniques algorithm.
K-means clustering technique is a centroid-based or distance-based algorithm that uses distances to assign a point to a cluster. Each cluster is associated with a centroid in this clustering technique. The main goal of this technique is to maximize the sum of distances between the points and their respective cluster centroid. The steps involve selecting the number of clusters (k), randomly selecting the centroid for each cluster, assigning each point to the closest cluster centroid, computing the centroids of newly formed clusters, and then repeating the process of assigning each point to the closest cluster centroid to form another new cluster.
Hierarchical Clustering is a popular clustering technique for estimating patterns in multi-dimensional data-sets. Examining groups of data with a similar pattern can lead to a better understanding of the functions and state of the data set used in this thesis. The steps involved are to determine the distance between each data point, consider all of the data points to be individual clusters, determine the number of cluster groups required, calculate the proximity of the new clusters and merge the two closest clusters to form new clusters and finally, merge all available clusters to form a new cluster. Ward's method was used in the hierarchical clustering method to determine the similarity between two clusters. This method was chosen due to its excellent performance in separating clusters even when there is noise between clusters. The ward's linkage method was used to compute the distance matrix between the closest features considered in the hierarchical clustering technique [89]. It can be expressed mathematically using Equation (2.16).

$$D(c_1, c_2) = \frac{\|c_1\|\|c_2\|}{\|c_1\| + \|c_2\|}\|c_1 - c_2\|^2, \tag{2.16}$$

where c_1 and c_2 are the two closest clusters and the norm is defined on the set of subsets of the data set (e.g., the norm of their barycenter).

Table 2.3: Hierarchical clustering algorithm parameters.

Hyperparameter	**Nature**
Number of clusters	Three
Affinity	Euclidean
Linkage	Ward

A Dendrogram was used to depict the hierarchical clustering method. A dendrogram represents the results of cluster analysis; it reveals all of the steps taken in the hierarchical algorithm, including the distance at which clusters merged [90]. Dendrograms provide a convenient method for exploring various data-combination possibilities. The number of clusters, affinity, and linkage are the parameters used in the training of a hierarchical algorithm, as shown in Table 2.3. All of these parameters are optimized to improve model optimization. The number of clusters to form and centroids to generate was set at three. The affinity used for this technique was Euclidean distance, which refers to the distance between all points in the selected data. The ward linkage method estimates the proximity between two clusters.

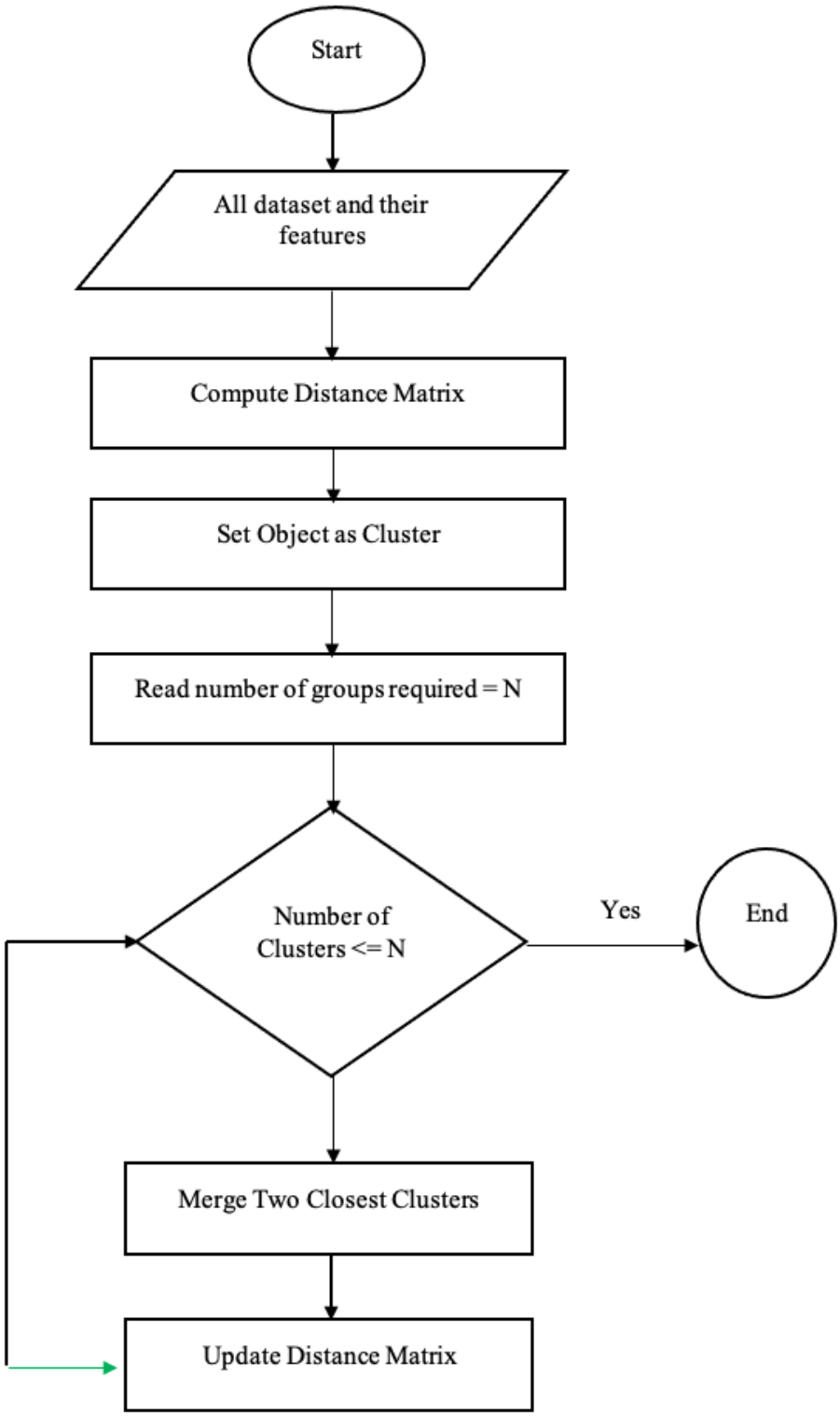


Figure 2.8: Flowchart of clustering technique.

2.2.4 Deep Structured Learning Modeling

Deep structured learning modeling involves the use of artificial intelligence or artificial neural network in modeling different phenomena; however, in this case it will be used to model COVID-19 outbreak from several of its epidemiological, socio-economic and demographic determinants. Artificial neural network modeling was developed to study interconnected networks which can be in form of nodes or neurons with many layers involved. It can be applied to biological systems, stochastic processes, computing among several others. It is divided into two major classes which are feed-forward artificial neural network and recurrent artificial neural networks. These classes are based on several improvements over the years which show that one model is performed better than the others when applied to an experimental work. In Table 2.4, the list of hyper-parameters for artificial neural network modeling, their definitions and values are presented while in Figure 2.11, the work flow of how the deep structure learning modeling was carried out using France, India and Turkey as an example in the flow chart.

Recurrent artificial neural networks

Recurrent artificial neural network is a type of artificial neural network in which node connections form a directed or not directed graph along a temporal sequence. This enables it to exhibit temporal dynamic behavior [91, 92]. There are two major types of recurrent artificial neural networks which are Long Short-Term Neural Network (LSTM) and Gated Recurrent Unit (GRU). The most advanced types of Recurrent Neural Networks are LSTM and GRU (RNN). The LSTM addresses an inherent problem of recurrent neural networks, the gradient vanishing problem. It is intended to model long-term dependencies and determine the optimal time lag for the time series by granting the memory unit the ability to decide, remember, or forget some information. This complex positioning among the recurrent neural networks allows the LSTM network to recall past data, making it easier to create connections between current and past data points, allowing the network to find patterns that play out over time. The GRU is similar to the LSTM in that it uses gates to control data set flow, but unlike the LSTM, it lacks a separate cell state and instead has only a hidden state. The GRU has some parameters that help to accelerate data set training. Figure 2.9 depicts the LSTM and GRU architectures.

Feed-forward artificial neural networks

Feed-forward artificial neural network is sometimes called deep feedforward artificial neural network are models that enable information to flow through the function being evaluated, the intermediate computations used to define the function, and finally give an output. There are no feedback connections where the model's outputs are fed back into itself. There are basically four types of this model which are Convolutional Neural Network (CNN), Deep Neural Network (DNN), Extreme Learning Machine (ELM) and Multilayer Perceptron (MLP).
The hidden layer of an artificial neural network is used to store and evaluate how significant one of its inputs is for the output. Figure 3b depicts an artificial neural network with only one hidden layer. A CNN (see Figure 2.10(a)) is a type of neural network composed of neurons with learnable weights and biases and it requires significantly less pre-processing than other classification algorithms. It is made up of filters that are applied to the input data set, effectively condensing the data set into a smaller resolution. CNN is well-positioned

to remove noise in the inputed data set so that it can be fed into basic neural networks. The advantage of this method is that it approaches the data set spatially [93].
A DNN is a multi-layer artificial neural network in which data flows from the input layer to the output layer without going backward, and the links between the layers only go forward and never touch another node again. It is also referred to as a feed forward network. A DNN is shown in Figure 2.10(b). A MLP is a feed-forward ANN with at least three layers of nodes: an input layer, a hidden layer, and an output layer, as shown in Figure 2.10(b), whereas ELM is used in the training algorithm for a feed-forward artificial neural network, which converges much faster than traditional methods and yields promising results [94].

Neural prophet forecasting modeling

Neural prophet forecasting model is an extension of Equation (2.9) with neural network components (auto-regression and covariates modules) and it is as expressed as follows:

$$y(t) = g(t) + s(t) + h(t) + f(t) + a(t) + l(t) + \varepsilon(t). \tag{2.17}$$

For all $m \in [1, h], h$ is the step length, and we have:

$$y(t+m-1) = g(t+m-1)+s(t+m-1)+h(t+m-1)+f(t+m-1)+a(t+m-1)+l(t+m-1),$$

where $f(t)$ is the forecast covariates or future known variables, $a(t)$ auto-regression of the form Equation (2.1) and $l(t)$ is the lagging current covariate by m time period or lagged observations of the known variable. Other parameters are defined in the same way as before. $f(t)$ can be multiplicative like seasonality which was earlier defined with a trainable coefficient σ_m then it can be defined as follows:

$$f(t) = \sum f_m^*(t)$$

where, for each m:

$$f_m(t) = \sigma_m f(t)$$

$$f_m^*(t) = \begin{cases} g(t) \times f_m(t) \text{ if multiplicative} \\ f_m(t), \text{ otherwise.} \end{cases}$$

Given a set of covariates x and if we create a separate lagged regressor module for each of m covariates x of length T, $l(t)$ can be expressed as follows:

$$l(t) = \sum_p l_x(x_{t-1}, \ldots, x_{t-p}).$$

The p last observations of the covariate x are the inputs to the module (instead of the series y itself as in auto-regressor) [95].

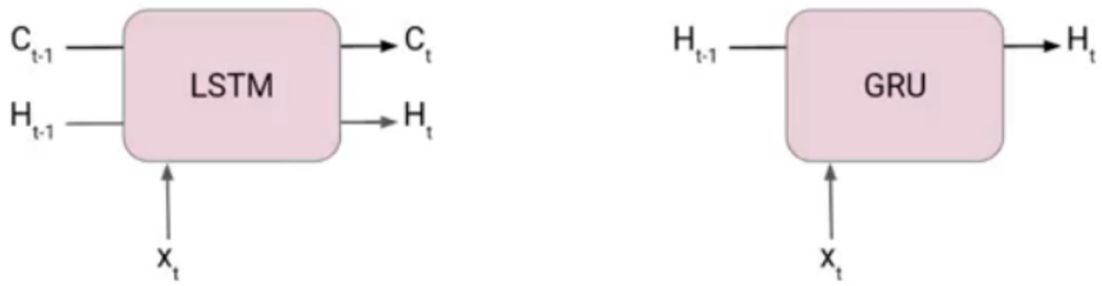


Figure 2.9: LSTM and GRU architectures (after ([96]). The calculation unit uses cell state (C) and hidden state (H) variables for LSTM and only H for GRU, with a control variable (X_t) for both.

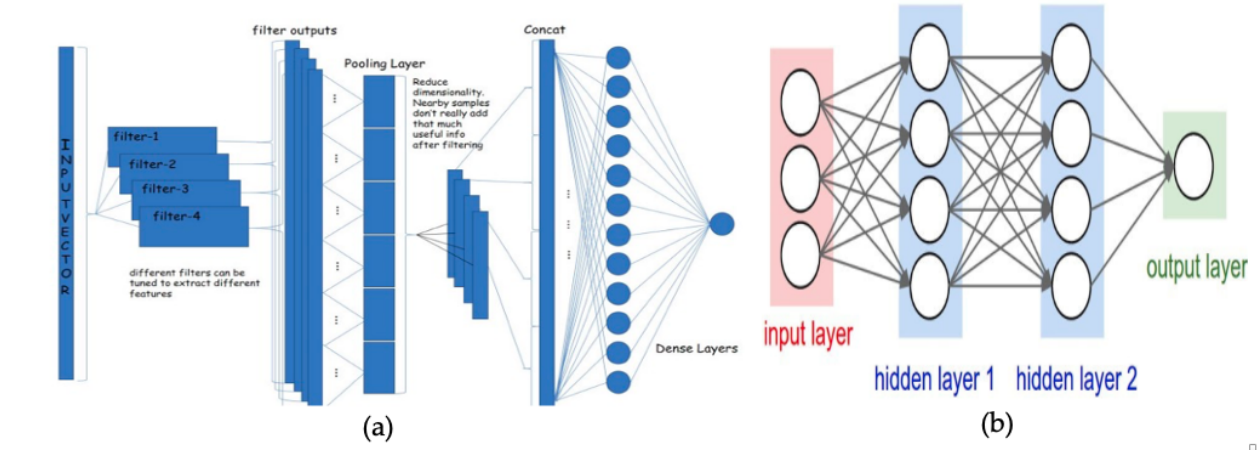


Figure 2.10: (a) A CNN sequence illustration (after [96]). (b) Deep neural network with two or more hidden layers.

Table 2.4: Artificial neural network models parameters.

Hyperparameter	**Definition**	**Values**
Batch size and max-pooling layers	Batch size is the number of sub-samples given to the neural networks after which parameters is updated	30, 32, 64, 128, 256, ...
Number of epoch	It is the number of iterations or times in training a neural network	100, ...
Number of hidden layers and units	It is the layers between input layers and output layers	32, 50, 64, 128, 256, ...
Dropout	It is used to regularize artificial neural network models to avoid over-fitting	20 - 50%
Optimizer	It is used to change the attributes of the neural networks such as weights and learning rate so as to reduce losses	Adam, stochastic gradient and gradient descent
Activation function	It decides whether the neurons should be activated or not	Sigmoid, Rectified Linear Unit (ReLU) and tanh
Filter size/Kernel in CNN layers	It helps to pad the input volume with zeros in such a way that the CNN layers does not alter the dimension of the input	3×3 or 5×5
Pooling size	It reduces the dimension of the feature map	2×2
Verbose	It is a setting that allows you to see wordy information in neural network modeling while it is training.	0, 0.2, 1, 2, ...
Validation split	It's a float value for validation data which is not used for the training, but to evaluate the loss and the accuracy	Between 0 and 1

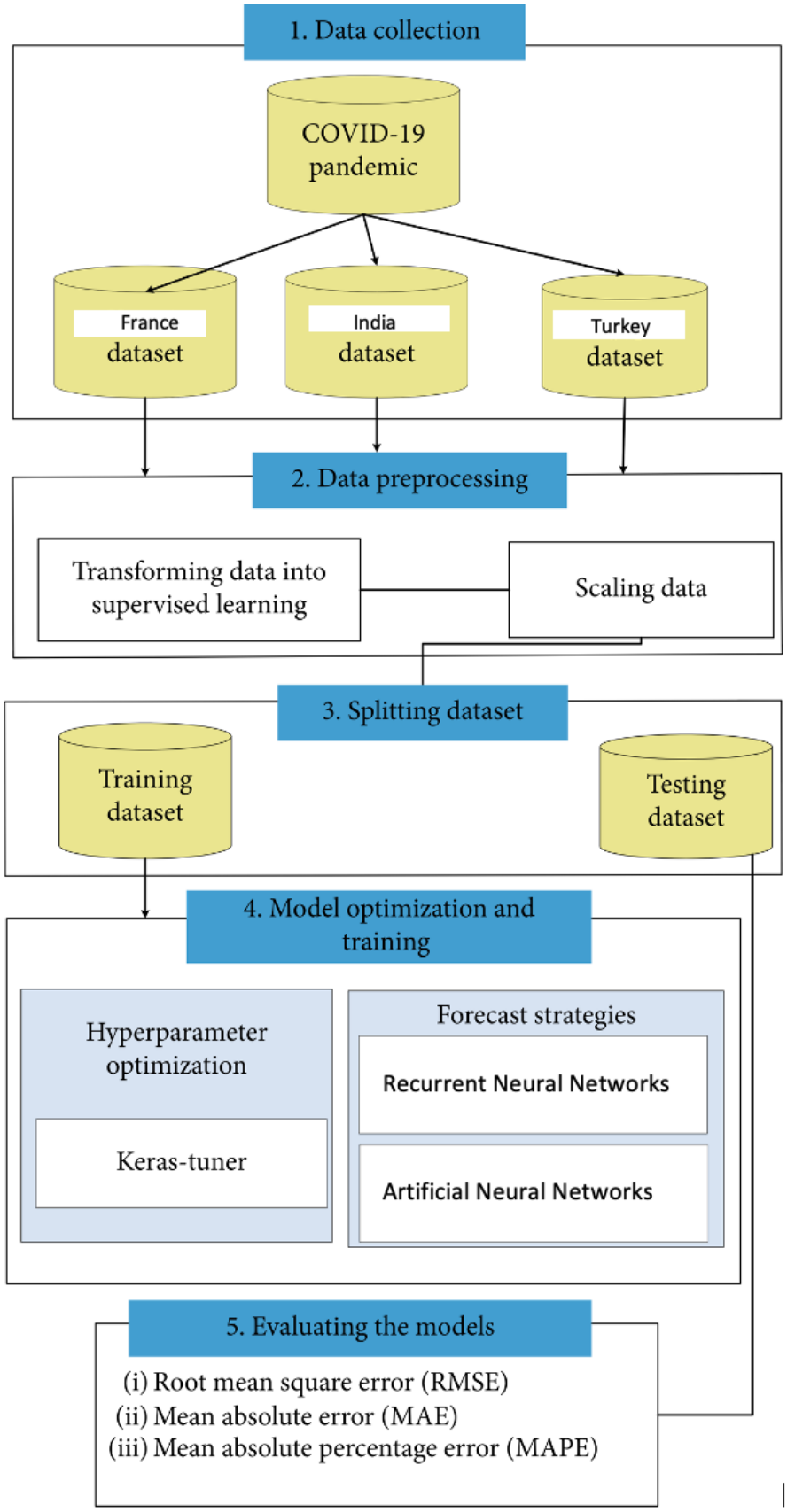


Figure 2.11: Flowchart of deep structured learning modeling.

2.2.5 Functional Data Analysis (FDA)

Functional data analysis deals with data from curves, surfaces, or anything else that varies along a continuum (e.g. time, spatial location, wavelength, probability, etc.). Data is only sampled at certain points along the continuum, either very precisely or with intricate noise structure, because the observed data are generally filled with observational errors (or noise) that are superimposed on the underlying signal. The main assumption is that the underlying random processes are intrinsically regular. In the real world, we could see a scenario with M processes being observed at the same time [97]. Let t be a one-dimensional argument sometimes referred as time. Functions of t are observed over a discrete grid $\{t_1, \ldots, t_K\} \in \Delta$ at sampling values t_k, which may or may not be equally spaced. Let Y be a vector of M functional data $Y = [Y_1^T, \ldots, Y_M^T]^T$, each functional data is expressed as follows:

$$Y_{ik} = X_i(t_k) + \varepsilon_{ik}, 1 \leq k \leq K, 1 \leq i \leq M,$$

where Y_{ik} is a noisy observation of the stochastic process $X_i(t_k)$ and ε_{ik} is a random error with zero mean and variance function σ_i^2 associated with the i^{th} functional raw data. In vector notation, the $X_i(t)$ is of the form:

$$X_i(t) \approx c_i^T.\alpha(t), \forall t \in \Delta, i = 1, \ldots, M,$$

where c_i and $\alpha(t)$ are n-vectors.

Smoothing techniques

The idea behind smoothing procedures is to estimate a curve that fits the data set well while not becoming too 'wiggly' or locally variable. This is accomplished by allowing the curve estimate at a point t to be heavily influenced by observations near t. B-spline technique is one of the tools used in smoothing (curve estimation modeling) a functional data and this can be done by changing the number of basis functions ($n = 2, 3, 4, \ldots, p$). The most commonly used basis functions are Fourier, polynomial, splines and wavelets. A mathematical expression for the basis representation for the curve of the functional data set is of the form:

$$f(t) = \sum_{j=1}^{n} b_j \alpha_j(t), \forall\ t \in \Delta, \tag{2.18}$$

where b_j's are basic coefficients and α_j's are the basis functions for $j = 1, \ldots, n$.
Sometimes one can use the Fourier basis for the functions to further see the variations in the curves, which is the most appropriate basis for periodic functions defined on an interval Δ where α_j's take the following form:

$$\alpha_0(t) = 1/\sqrt{|\Delta|}, \alpha_{2r-1}(t) = \frac{\sin(r\omega t)}{\sqrt{|\Delta|}/2} \text{ and } \alpha_{2r}(t) = \frac{\cos(r\omega t)}{\sqrt{|\Delta|}/2},$$

for $r = 1, \ldots, \frac{n-1}{2}$, where n is the number of basis functions and n must be odd to compute a Fourier basis. The frequency ω determines the period and the length of the interval $|\Delta| = 2\pi/\omega$.
The non-parametric method (kernel smoothing) of smoothing for functional data is based on the smoothing matrix M given:

$$m_{ij} = \frac{1}{h} K(\frac{t_i - t_j}{h}), \tag{2.19}$$

$$m(h) = (s_j(t_i)) = \frac{K(\frac{t_i - t_j}{h})}{\sum_{k=1}^{T} K(\frac{t_k - t_j}{h})}, \tag{2.20}$$

where $K(.)$ is the Kernel function and h is smoothing parameter.

Functional principal component analysis

Principal component analysis (PCA) is a dimension reduction analysis tool in unsupervised machine learning modeling, whereas functional principal component analysis (FPCA) is a dimension reduction with high correlation in functional data analysis that completes the statistical tools used in biomedical data modeling, particularly epidemiologic data.
Let $\{x_i(t)\}_{i=1,m}$ be a given set of function and let α be a weight. To compute functional principal components, the following steps are involved:
- We first find the principal component weight function $\alpha_1(t)$ for which the principal component score is given as follows:

$$f_{i1} = \int \alpha_1(t) x_i(t) dt$$

while maximizing $\sum_{i=1,m} f_{i1}^2$ is subjected to:

$$\int \alpha_1^2(t) dt = \|\alpha_1\|^2 = 1$$

- Next, the weight function $\alpha_2(t)$ is computed and the principal component score maximizing $\sum_{i=1,m} f_{i2}^2$ is subjected to $\|\alpha_2\|^2 = 1$ and to the additional constraint:

$$\int \alpha_2(t) \alpha_1(t) dt = 0$$

- Then, the process is repeated for as many iterations.

Functional clustering analysis

The two-stage approaches for functional data clustering consist of a first step, referred to as the filtering step, in which the dimension of data is reduced, and a second step in which classical clustering tools for finite dimensional data are used. The reducing dimension step consists in general in approximating the curves to a finite basis of functions. Because of their superior properties, spline bases are a popular choice. Clustering algorithms, such as the k-means algorithm, are commonly used to estimate clusters of functional data. Nonparametric approaches for functional data clustering include hierarchical and k-means clustering, which are tools that use standard nonparametric clustering techniques with specific distances or dissimilarities. Several works [98, 99] consider the following measures of proximity between two curves x_i and x_j:

$$d_l(x_i, x_j) = \left(\int_T (x_i^l(t) - x_j^l(t))^2 dt \right)^{1/2},$$

where x^l is the $l - th$ derivative of x.
It has been proposed [98, 99, 100] to use hierarchical clustering in conjunction with the distance d_0 (the L_2-metric) or d_1 (the semi-norm in the Hilbert space) or with the semi-metric d_2. Another proposal is the use of k-means algorithm with d_0, d_1 and with $(d_0^2 + d_1^2)^{1/2}$, as well

as d_0 with k-means for Gaussian processes. It was demonstrated, in particular, that cluster centers are linear combinations of functional principal component analysis eigenfunctions. The same distance d_0 with k-means is considered when defining time-dependent clustering. Following the approach used to estimate the distance d_0, nonparametric analysis can be applied to raw-data clustering or to two-stage analysis. Indeed, if d_0 is approximated directly using discrete observations of curves, nonparametric analysis is equivalent to raw-data clustering analysis. Similarly, if an approximation of the curves into a finite basis is used to approximate d_0, nonparametric analysis is equivalent to two-stage methods with the same basis approximation.

Functional linear regression analysis

This section describes two functional linear regression analysis methodologies which are (i) function to scalar linear regression and (ii) function-on-function linear regression.
To describe the function to scalar linear regression model, Let

$$Y =< \theta, X > +\varepsilon, \tag{2.21}$$

where θ is the unknown function of the model, X is a functional covariate belonging to some functional space $\mathbb{H}$ which is endowed with an inner product $< \cdot, \cdot >, Y$ is the response variable and ε is the random error term. Usually, $\mathbb{H}$ is $L^2([a, b])$ of square integrable functions on some real compact interval $[a, b]$ and

$$< f, g >= \int_a^b f(t)g(t)dt, \tag{2.22}$$

is the corresponding inner product, where the functions $f, g \in L^2([a, b])$. Considering $C = [0, 1]$, so the Equation (2.21) can be written as:

$$Y = \int_0^1 \theta(t)X(t)dt + \varepsilon, \tag{2.23}$$

where θ is a square integrable function which is defined on C and ε is a random variable such that $\mathbb{E}(\varepsilon) = 0$ and $\mathbb{E}(X(t)\varepsilon) = 0$, then Equation (2.23) can be rewritten as:

$$Y = (\Psi X) + \varepsilon, \tag{2.24}$$

where Ψ represents the integral.
If the functional linear regression model is considered as a functional input/output regression model where $y(t)$ is response variable and $x(t)$'s are predictors at each time t, i.e., $x(t) \to y(t)$. The functional linear model with an intercept is of the form:

$$y(t) = \beta_0(t) + \int \beta(s, t)x(s)ds + \varepsilon. \tag{2.25}$$

Chapter 3

Mathematical Modeling of COVID-19 Outbreak

This chapter deals with a wide range of mathematical models developed to model the COVID-19 pandemic. Section 3.2 describes the use of matrix algebra to estimate the daily reproduction numbers during the COVID-19 outbreak. The discrete dynamics of the epidemic's growth equation is presented with the estimation of the daily reproduction numbers using a de-convolution technique on a series of new COVID-19 cases, then the discussion of the role of noise on the stability of the epidemic's dynamics using a stochastic framework is proposed. The shapes of the daily reproduction number distribution in the COVID-19 outbreak dynamics are similar to other viral diseases, such as influenza, with two successive reactions from two immune defense barriers, innate cellular immunity first, which is insufficient if symptoms persist, then adaptive immunity, resulting in a transient decrease in contagiousness between the two phases. In Section 3.3, the estimation of maximum reproduction number using continuous formulation of the epidemic's growth equation is presented and applied to different countries in order to show the effectiveness of the method. In Section 3.4 the identification of the first point of inflexion using Bernoulli SI equation and the estimation of transmission rate of the COVID-19 outbreak is presented. The method is applied to real data from Cameroon and its regions, then to the world data. The results derived from this analysis shows its applicability to countries whose epidemic curves did not reach the turning point. In Section 3.5 a discrete epidemic model on the ageing and lifespan loss in COVID-19 outbreak and the role of comorbidities of infected patients using the Usher's model for different age classes and gender (male and female) at sub-population level is presented. Also, the proportion of sensitivity of the model parameters so as to see the most important parameter at different age groups is established. The result is consistent with the observed cases and the semi-quantitative mathematical modeling approach at the population level. The Section 3.6 describes a mathematical age dependent model for two age groups (20-64 years and people above 64) named Susceptible-Infectious-Goneanewsusceptible-Recovered (SIGR) model for COVID-19 outbreak and then some mathematical results by showing the positivity, boundedness, stability, existence and uniqueness of the solution are proved. Numerical simulations of the model with parameters coming from Kuwait, France and Cameroon are presented. The results are centred around the role of different parameters used in the model mainly the effect of vaccination on the epidemic dynamics. Finally, Section 3.7 discusses the perspectives of this chapter and some conclusions based on the models developed.

3.1 Introduction

Mathematical modeling of infectious diseases is important to better understand the evolution of any pandemic most especially now that the world is faced with the COVID-19 outbreak which was first discovered in Wuhan, China in December, 2019 and since then we have had different waves of the pandemic as most countries is battling with their fifth wave due to the most recent variant, the omicron variant, which is known for its high contagiousness but not as deadly as previous variant Delta. Recently, in Ireland and some other countries, we are having a rise in hospitalization cases and daily new cases among children, primary school children, their adult contacts and their parents which is one of the reason why the vaccination campaign is now extended to younger ones.
COVID-19 pandemics has affected every facet of our lives. The introduction of various mitigation measures has helped to curtail the spread of the disease but this has had adverse effects on the well being of many people due to isolation, psychological effect, lost of jobs and many more impacts on the people. The patterns of cases have shown that the unvaccinated are at far greater risk and there is a lot of debates on declaring that COVID-19 is endemic since the managing of the spread of the disease is getting difficult by the day due to an extended period.
As the virus keeps mutating, the onus lies on us as researchers to keep studying the patterns and the spread of the virus hence, the reason for which we propose different mathematical approaches to better understand the basic reproduction number, the transmission rate from the first point of inflection and the dynamics of the spread of COVID-19 virus at the sub-population (age group) level.

3.2 Estimation of Daily Reproduction Rates in COVID-19 Outbreak

3.2.1 The basic reproduction number

The ability of an infectious disease to spread is quantified by the basic reproduction number R_0 (also called the average reproductive rate), a classical epidemiologic parameter that describes the transmissibility of an infectious disease and is equal to the number of susceptible individuals that an infectious individual can transmit the disease to during his contagiousness period.
Given an SI Equation of the form:

$$\begin{aligned} \frac{dS}{dt} &= -\nu SI, \\ \frac{dI}{dt} &= \nu SI - \mu I, \end{aligned} \tag{3.1}$$

where S and I are respectively the size of the susceptible and infectious populations. μ is the death rate and ν is the rate of contact. If we assume that S is approximately 1 i.e., everyone in the population is susceptible to the disease, then $dI/dt = \nu I - \mu I$. An individual is expected to infect a number of ν/μ secondary cases which represents the basic reproduction number R_0 at the individual level modeling i.e., $R_0 = \nu/\mu$. It has been established that if

$R_0 = \nu/\mu > 1$, the pandemic is an epidemic, and if $R_0 = \nu/\mu < 1$, the pandemic is extinct or the disease can be control but a times that is not the case.
Following the outbreak of the COVID-19 disease caused by coronavirus SARS CoV-2 which is the third coronavirus outbreak to occur in the past two decades. SARS CoV-2 can be transmitted from person to person by respiratory droplets and through contact and fomites. Therefore, the severity of the disease symptoms, such as cough and sputum (see Figure 3.2), and their viral load, are often the most important factors in the virus's ability to spread, and these factors can change rapidly within only a few days during an individual's period of contagiousness. For contagious diseases like COVID-19, the transmissibility is not a biological constant: it is affected by numerous factors, including endogenous factors, such as the concentration of the virus in aerosols emitted by the patient (variable during his contagiousness period), and exogenous factors, such as geo-climatic, demographic, socio-behavioral and economic factors governing pathogen transmission (variable during the outbreak's history) [101, 102, 103] which will be discussed and taking into consideration in the modeling of the COVID-19 disease in some part of this thesis.
In epidemiology, there are essentially two broad ways to calculate R_0, which correspond to the individual-level modeling and to the population-level modeling. At the individual level, if we suppose the susceptible population size constant (hypothesis valid during the exponential phase of an epidemic), the daily reproduction rates of an individual are typically non-constant over his contagiousness period, and the calculations we present define a new method for estimating R_0, as the sum of the daily reproduction rates. This new approach allows us to have a clearer view on the respective influence on the transmission rate by endogenous factors (depending on the level of immunologic defenses of an individual) and exogenous factors (depending on environmental conditions). Due to these exogenous factors, R_0 might change seasonally, but these factor variations are not significant if a very short period of time is considered. R_0 depends also on endogenous factors such as the viral load [104] of the infectious individuals during their contagiousness period, and the variations in this viral load (see Figure 3.2) must be considered in both theoretical and applied studies on the COVID-19 outbreak, in which the authors estimate a unique reproduction number R_0 linked to the Malthusian growth parameter of the exponential phase of the epidemic, during which R_0 is greater than 1 (Figure 3.1). The corresponding model has been examined in depth, because it is useful and important for various applications, but the distribution of the daily reproduction number R_j at day j of an individual's contagiousness period is rarely considered within a stochastic framework.
We therefore defined a partial reproduction number for each day of an individual's contagiousness period, and, assuming initially that this number was the same for all individuals, we obtained the evolution equation for the number of new daily cases in a population. Assuming that the distribution of partial reproduction numbers (referred to as daily for simplicity) was subject to fluctuations, we calculated the consequences for their estimation, and we estimated them for a large number of countries, taking a duration of contagiousness of 3 followed by 6 and 7 days.
When this distribution is considered, it is possible to calculate its entropy as a parameter quantifying its uniformity and to simulate the dynamics of the infectious disease either using a Markovian model such as that defined in Delbrück's approach [105] or a classical discrete or ODE SIR deterministic model. In the Markovian case, R_0 can be calculated from the evolutionary entropy defined by L. Demetrius as the Kolmogorov–Sinaï entropy of the corresponding random process [106], which measures the stability of the invariant measure,

dividing the population into the subpopulations S (individuals susceptible to but not yet infected with the disease), I (infectious individuals) and R (individuals who have recovered from the disease and now have immunity to it). In the deterministic case, R_0 corresponds to the Malthusian parameter quantifying its exponential growth, and the stability of the asymptotic steady state depends on the subdominant eigenvalue [107].

3.2.2 The contagion mechanism from a first infectious case zero

In [108], Eric Goles and his colleagues investigated the spread of an epidemics where an infected patient can recover only if he is not isolated from a central "node" (agent or set of agents), which could represent a public health or medical authority able to prescribe mitigation measures, treatments or vaccines. They used a SIS (for Susceptible-Infected-Susceptible) model, on an underlying network of individuals susceptible to enter in contact or to have interactions implying a physical proximity. Then, the infected individuals contaminate healthy people, and the epidemic spreads. The infected people can only recover when it remains connected to the central node through a path of contacts involving only healthy individuals. In such an epidemic, clusters of infected people will contaminate their healthy strict interior (with empty intersection with cluster boundary) because it has no contact with the central node no path. For example, in an Apollonian network (i.e., an undirected graph formed by a process of recursively subdividing a triangle into three smaller triangles) an infection without reinfection (i.e., with a perfect immunization) starting at the central nodes of such a network will spread until the periphery and stops; on the contrary, an infection starting on the network boundary will reach the central node, or be blocked if there is simultaneously a start in different central nodes (Figure 3.1 Top left). A more realistic network, called the "school network" authorizes long distance links breaking the regular vertices topology of an ideal Apollonian network and seems to be more realistic for modelling interaction networks in which a real epidemic diffuses [109].
Let us suppose that the secondary infected individuals are recruited from the centre of the sphere of influence of an infectious case zero and that the next infected individuals remain on a sphere centred on case 0, by just widening its radius on day 2. Therefore, the susceptible individuals $C(j)$, which each infectious on day $j-1$ can recruit, are on a part of the sphere of influence of case 0 reached at day j (rectangles on Figure 3.1).
Let us suppose now that the daily reproduction number denoted as a depends on the day $j : a_j => C(j)$, where $C(j)$ represents the number of susceptible individuals, which can be met by one contagious individual at day j. If infected individuals (supposed to all be contagious) at day j are denoted by I_j, we have:

$$X_j = \frac{\Delta I_j}{\Delta j} = \frac{I_{j+1} - I_j}{j+1-j} = \nu C(j) I_j \tag{3.2}$$

Let us suppose that the first infectious individual 0 recruits from the centre of its sphere of influence secondary infected individuals remaining in this sphere, and that the susceptible individuals recruited by the I_j infectious individuals present at day j are located on a part of the sphere of centered on the first infectious 0 obtained by widening its radius (Figure 3.1). Then, we can consider that the function $C(j)$ increases, then saturates due to the fact that an infectious individual can meet only a limited number of susceptible individuals as the sphere grows. We can propose for $C(j)$ the functional form $C(j) = S(j)/(c + S(j))$, where $S(j)$ is the number of susceptible individuals at day j. Then, we can write the following equation,

taking into account the mortality rate μ:

$$X_j = \frac{\Delta I_j}{\Delta j} = \nu C(j) I_j - \mu I_j = \nu I_j \frac{S(j)}{c + S(j)} - \mu I_j \tag{3.3}$$

This discrete version is used much less than the classic continuous version, corresponding to the ODE SIR model, with which we will show a natural link. Indeed, the discrete Equation (3.3) is close to SIR Equation (3.4) , if the value of c is greater than that of S:

$$\frac{dI}{dt} = \nu \frac{IS}{(c + S)} - \mu I \tag{3.4}$$

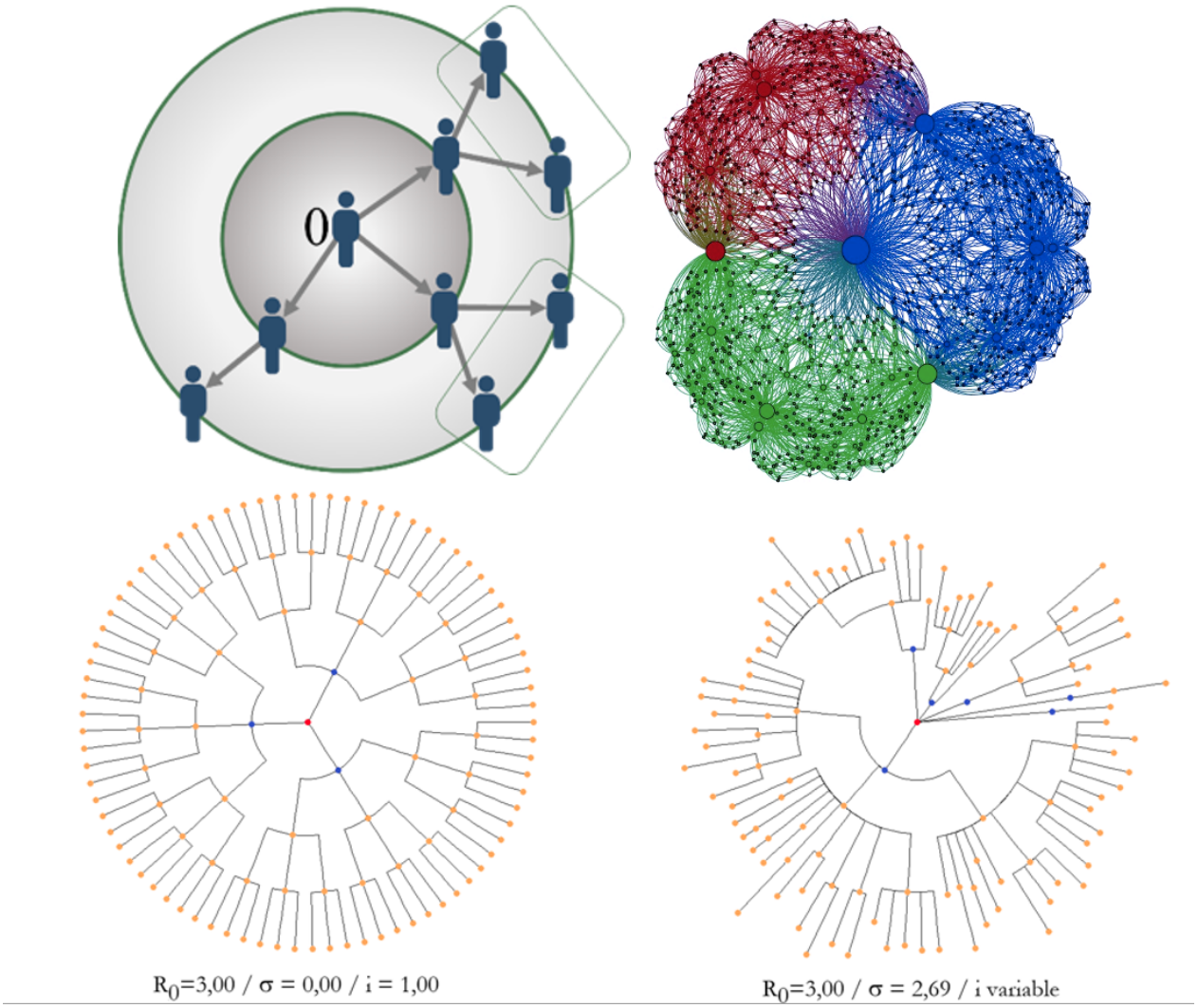


Figure 3.1: **Top left**: Spread of an epidemic from a first infectious case 0 (located at its influence sphere center) progressively infecting its neighbours in some regions (rectangles) on successive spheres. **Top right**: Apollonian interaction graph with 1096 nodes and average degree of connectivity k approximately 5.99 (after [108]). **Bottom left:** Spread of an epidemic from a first infectious "patient zero" (in red) located on the centre of its influence sphere made of the successive generations of infected individuals, with same value of the reproduction number $R_0 = 3$ and same inter-generation time interval i between two successive generations of infected. **Bottom right**: same epidemic, but with a stochastic dynamics defined by a uniform distribution of the reproduction number on an interval centred in R_0 with a standard deviation σ, and a random inter-generation time interval i (after [110]).

3.2.3 The biphasic pattern of the virulence curve of coronaviruses

Mostly, the clinical course of patients with seasonal influenza shows a biphasic occurrence of symptoms with two distinct peaks. Patients have a classic influenza disease followed by an improvement period and a recurrence of the symptoms [111]. The influenza RNA virus shedding (the time during which a person might be contagious to another person) increases sharply one half to one day after infection, peaks on day 2 and persists for an average total duration of 4.5 days, between 3 and 6 days, which explains why we will choose in the following contagiousness duration these extreme values, i.e., either 3 to 7 days, depending on the positivity of the estimated daily reproduction numbers. It is common to consider this biphasic evolution of influenza clinically: after incubation of one day, there is a high fever (39–40°C), then a drop in temperature before rising, hence the term "V" fever. The other symptoms, such as coughing, often also have this improvement on the second day of the flu attack: after a first feverish rise (39–39.5°C), the temperature drops to 38°C on the second day, then rises before disappearing on the 5th day, the fever being accompanied by respiratory signs (coughing, sneezing, clear rhinorrhea, etc.). By looking at the shape of virulence curves observed in coronavirus patients [110, 112, 113, 114, 115], we often see this biphasic pattern. Figure 3.2 shows a U-shaped evolution for the viral load in real [116] and in simulated [117] COVID-19 patients, and in real influenza-infected animals for the viral load and the body temperature [118].

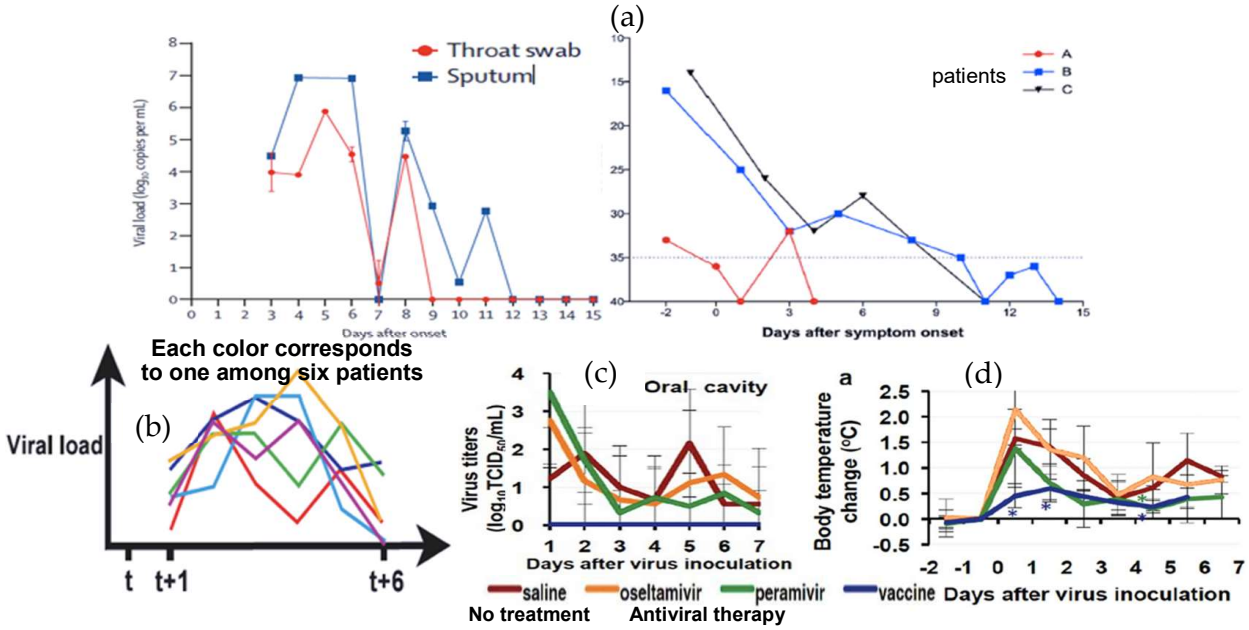


Figure 3.2: (**a**) Viral load in real COVID-19 patients [104], (**b**) in influenza-simulated patients [116] and (**c**) in real influenza-infected animals (red curve [117]), and (**d**) body temperature in real influenza-infected animals (red curve [117]).

3.2.4 Weekly patterns in daily infected cases

Daily new infected cases are highly affected by weekdays, such that case numbers are lowest at the start of the week and increase afterwards. This pattern is observed at the world level, as well as at the level of almost every single country or USA state. Hence, in order

to estimate biologically meaningful reproduction numbers, clean of weekly patterns due to administrative constraints, analyses have to be restricted to specific periods shorter than a week, or at rare occasions when patterns escape the administrative constraints. This weekly phenomenon occurs during exponential increase as well as decrease phases of the pandemic and during endemic periods in numbers of daily cases (Figure 3.3). In addition, the daily new infected case record is discontinuous for many countries/regions, which frequently publish, on Monday or Tuesday, a cumulative count for that day and the weekend days. For example, Sweden typically publishes only four numbers over one week, the one on Tuesday cumulating cases for Saturday, Sunday and the two first weekdays. Discontinuity in records further limits the availability of data enabling detailed analyses of daily reproduction numbers and can be considered as extreme weekday effects on new case records due to various administrative constraints.

We calculated Pearson correlation coefficients r between a running window of daily new case numbers of 20 consecutive days and a running window of identical duration with different intervals between the two running windows. These Pearson correlation coefficients r typically peak with a lag of seven days between the two running windows.

The mean of these correlations are for each day of the week from Tuesday (data making up for the weekend underestimation) to Monday: 0.571, 0.514 (0.081), 0.383 (0.00008), 0.347 (0.000003), 0.381 (0.000006), 0.468 (0.000444) and 0.558 (0.03916), with, in parentheses, the p-value of the one-tailed paired t-test showing that the correlation observed with running windows starting Tuesday are more than the others (see Table A.1 in Appendix A). This could reflect a biological phenomenon of seven infection days. However, examination of the frequency distributions of lags for r maxima reveals, besides the median lag at 7 days, local maxima for multiples of 7 (14, 21, 28, 35, etc.). About 50 percent of all local maxima in r involve lags that are multiples of seven (seven included).

This excludes a biological causation, except if data periodicity comes from an entrainment by the weekly "Zeitgeber" of census, near the duration of the contagiousness interval. We tried to control for weekdays using two methods, and combinations thereof. For the first method, we calculated z-scores for each weekday, considering the mean number of cases for each weekday, and subtracted that mean from the observed number for a day (Figure 3.3). This delta was then divided by the standard deviation of the number of cases for that weekday. The mean and standard variation are calculated across the whole period of study for each weekday.

The second method implies data smoothing using a running window of 5 consecutive days, where the mean number of new cases calculated across the five days is subtracted from the number of new cases observed for the third day. Hence, data for a given day are compared to a mean including two previous, and two later days (Figure 3.4).

We constructed two further datasets, where z-scores are applied in the first to data after smoothing from the second method and are applied in the second data after smoothing from the first method (not shown) (Figure 3.5 and Figure 3.6).

These four datasets from daily new cases database [119] transformed according to different methods and combinations thereof designed to control for weekday were analysed using the running window method. Despite attempts at controlling for weekday effects, the median lag was always seven days across all four transformed datasets, and local maxima in lag distributions were multiples of seven. After data transformations, about 50 percent of all local maxima were lags that are multiples of seven, including seven.

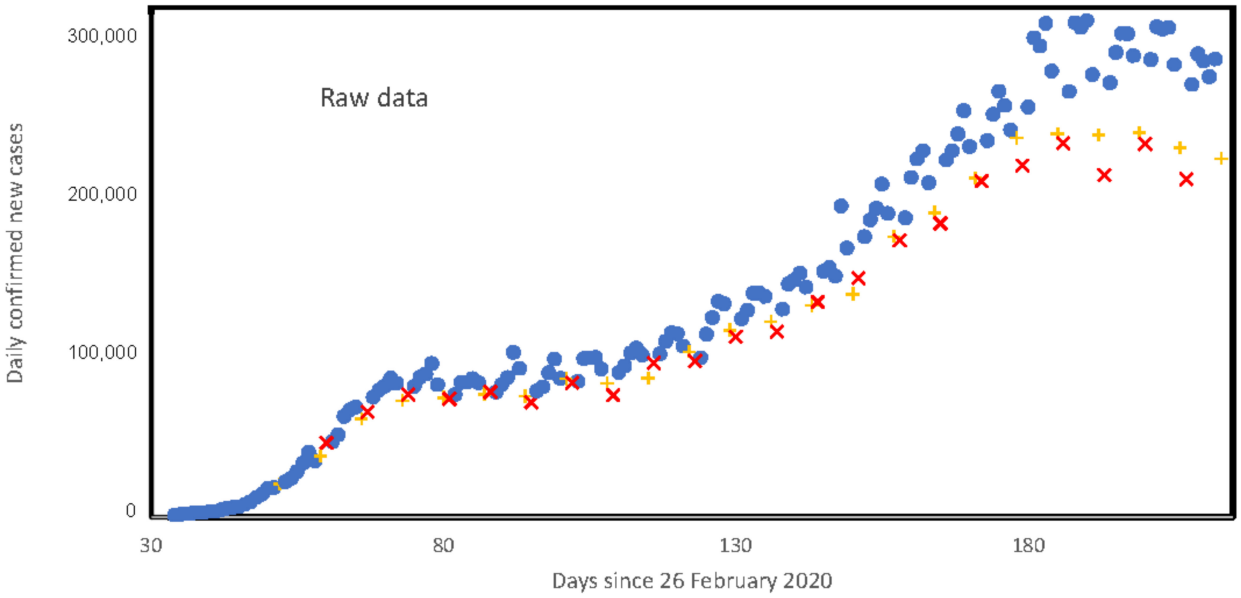


Figure 3.3: Confirmed world daily new cases (from [119]) as a function of days since 26 February until 23 August 2020 + indicates Sundays, x indicates Mondays.

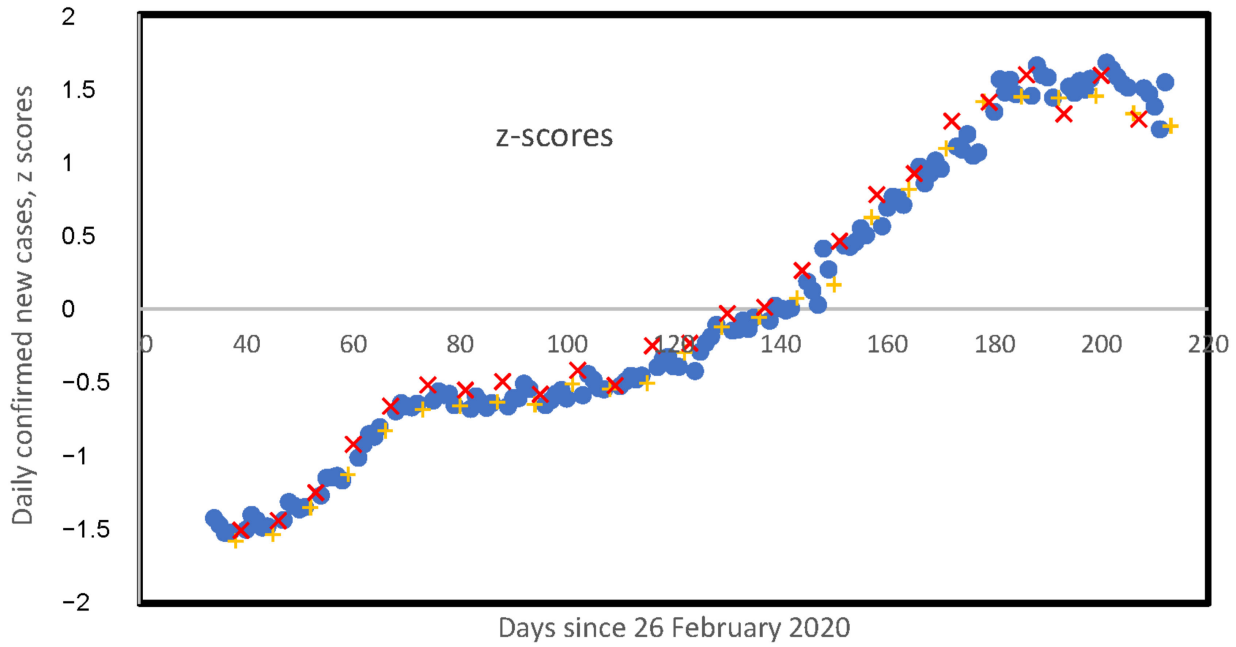


Figure 3.4: Z-transformed scores of confirmed world daily new cases [119], from Figure 3.3, as a function of days since 26 February 2020 until 23 August 2020 + indicates Sundays, x indicates Mondays. Z-transformations are specific to each weekday.

Visual inspection of plots of these transformed data versus time for daily new infected cases from the whole world shows systematic local biases in daily new infected cases (after transformation) on Sundays and Mondays, for all four transformed datasets, with Sundays and/or Mondays as local minima and/or local maxima, according to which method or combination thereof was applied to the data.

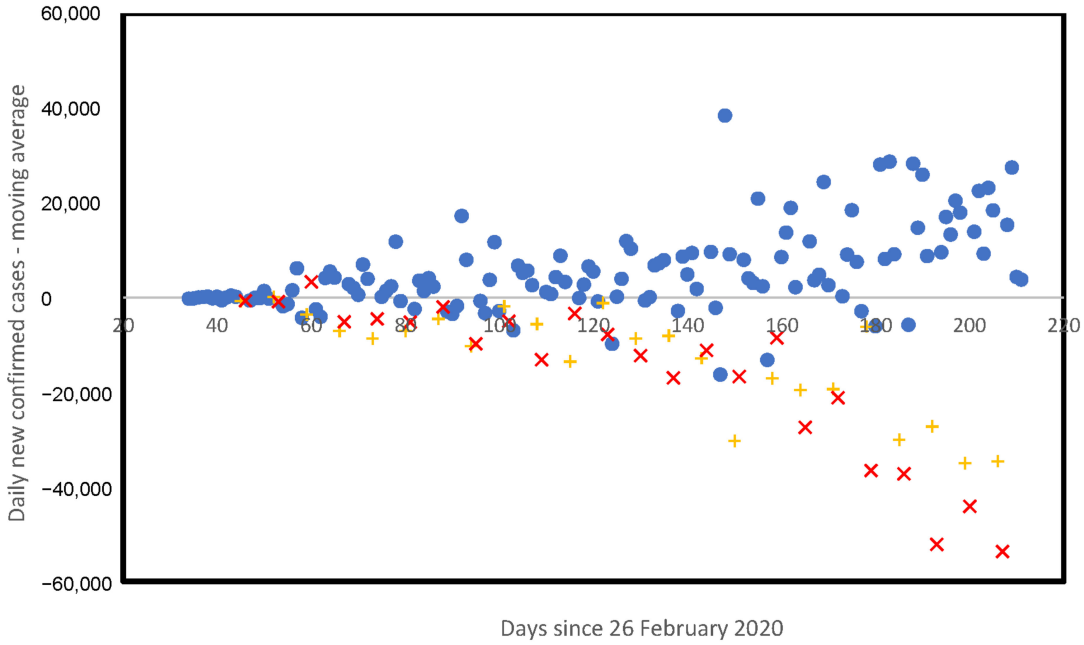


Figure 3.5: Smoothed confirmed world daily new cases [119], from (Figure 3.3, as a function of days since 26 February 2020 until 23 August 2020 + indicates Sundays, x indicates Mondays. For each specific day j, the mean number of confirmed daily new cases calculated for days $j-1, j-2, j, j+1$ and $j+2$ is subtracted from the number for day j.

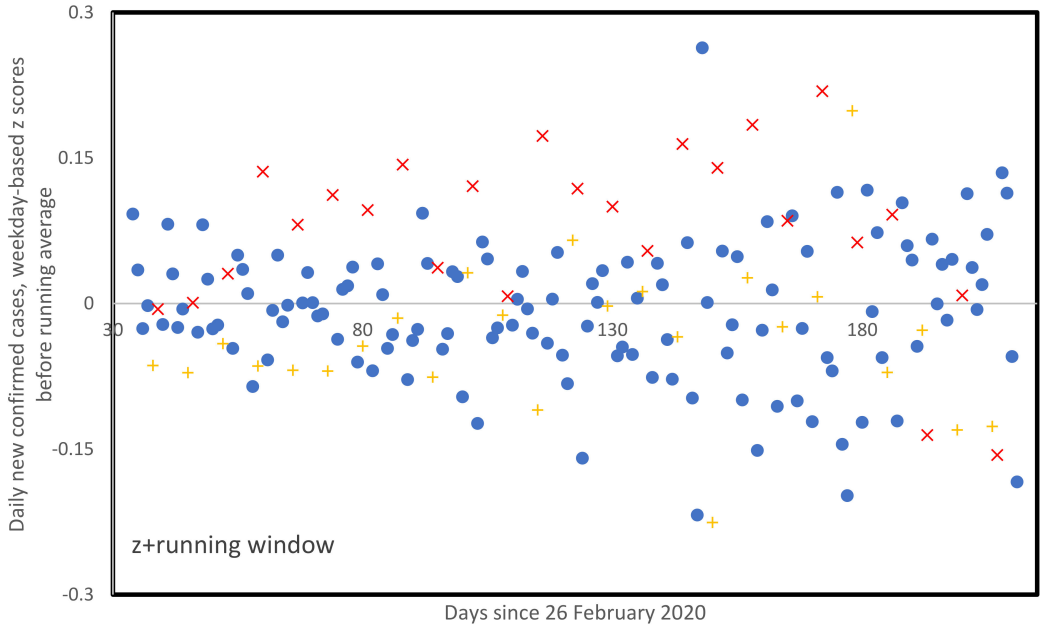


Figure 3.6: Smoothed confirmed world daily new cases [119] applied to z-scores from (Figure 3.4, as a function of days since 26 February 2020 until 23 August 2020 + indicates Sundays, x indicates Mondays. Z-transformations are specific to each weekday. For specific day j, the mean number of confirmed new cases calculated for days $j-1, j-2, j, j+1, j+2$ is subtracted from the number for day j.

Hence, the methods we used failed to neutralize the weekly patterns in daily new cases due to

administrative constraints. This issue highly limits the data available for detailed analyses of daily new cases aimed at estimating biologically relevant estimates of reproduction numbers at the level of short temporal scales. By smoothing on five consecutive days of raw data (confirmed world daily new infected cases [119]) and applying the z-transformation, we obtain a better result in Figure 3.7 than in Figure 3.6 in order to neutralize the weekly pattern. The reason is that the smoothing largely eliminates the counting defect during weekends due either to fewer hospital admissions and/or less systematic PCR tests or to a lack of staff at the end of the week to perform the counts.

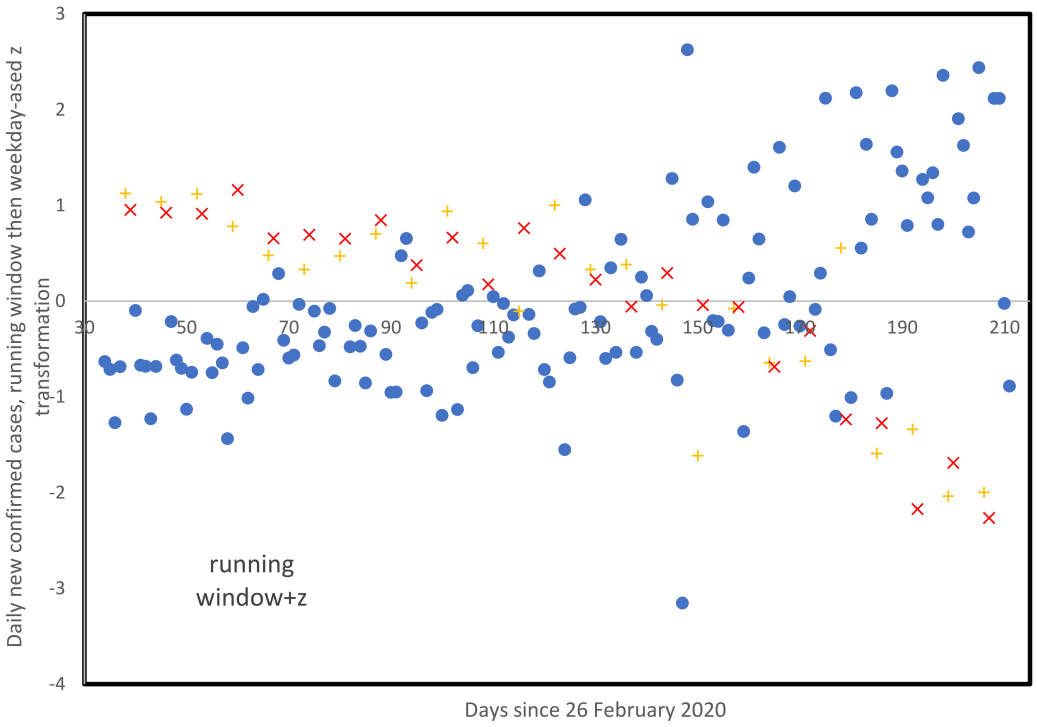


Figure 3.7: Z-transformed scores of smoothed confirmed world daily new cases [119] smoothed data from Figure 3.5, as a function of days since 26 February 2020 until 23 August 2020. + indicates Sundays, x indicates Mondays. Z-transformations are specific to each weekday.

3.2.5 A deterministic contagion discrete mechanism

In the following, we suppose that the susceptible population size remains constant, which constitutes a hypothesis valid during the exponential phase of epidemic waves. The ODE deterministic approach is linked by a common background consisting of the birth and death process approach used in the kinetics of molecular reactions by Delbrück [105], then in dynamical systems theory by numerous authors [106, 107], namely in modeling of the epidemic spread in exponential growth. In the ODE approach, the Malthusian parameter is the dominant eigenvalue.

The methodology chosen starts from an attempt to reconstruct an epidemic dynamic only from the observation of the number R_{ikj} of people infected at day j by a given infectious individual i during the k^{th} day of his period of contagiousness of length r. By summing on the number of new infectious individuals X_{j-k} at day $j-k$ where their contagiousness

started, the number of new infected people on day j is equal to:

$$X_j = \sum_{k=1,\dots,r} \sum_{i=1,\dots,X_{j-k}} R_{ikj} \tag{3.5}$$

We assume now that R_{ikj} is the same, equal to R_k, for all individuals i and day j, and depends only on day k. Then, we have:

$$X_j = \sum_{k=1,\dots,r} X_{j-k} R_k \tag{3.6}$$

$$X_{j-1} = \sum_{k=1,\dots,r} R_k X_{j-1-k}, \dots$$

$$X_{j-r+1} = \sum_{k=1,\dots,r} R_k X_{j-r-k}$$

The convolution Equation 3.6 is the basis of our modelling of the epidemic dynamics and the $\sum_{k=1,\dots,r} R_k$ plays the same role as the classical basic reproduction number R_0 [113].
Let us choose a simple deterministic rule of infection, and denote by X_j the number of new infected cases at day $j(j \geq 1)$, and $R_k(k = 1, \dots, r)$ the daily reproduction number at day k of the contagiousness period of length r for all infectious individuals. Then, we have obtained in Equation (3.6) by supposing that contagiousness behaviour is the same for all the infectious individuals: for example, if $r = 3$, for the number X_5 of new cases at day 5, the equation

$$X_5 = R_1 X_4 + R_2 X_3 + R_3 X_2$$

means that for example new cases at day 4 have contributed to new cases at day 5 with the term $R_1 X_4, R_1$ being the reproduction number at first day of contagiousness of new infected individuals at day 4.
In matrix form, we get:

$$X = MR \tag{3.7}$$

where $X = (X_j, \dots, X_{j-r-1})$ and $R = (R_1, \dots, R_r)$ are r-dimensional vectors and M is the $r - r$ matrix:

$$M = \begin{bmatrix} X_{j-1} & X_{j-2} & \cdots & X_{j-r} \\ X_{j-k} & X_{j-k-1} & \cdots & X_{j-k-r+1} \\ & & \cdots & \\ X_{j-r} & X_{j-r-1} & \cdots & X_{j-2r+1} \end{bmatrix} \tag{3.8}$$

It is easy to show that, if $X_0 = 1$ and $r = 5$ we obtain:

$$X_5 = R_1^5 + 4R_1^3 R_2 + 3R_1^2 R_3 + 3R_1 R_2^2 + 2R_2 R_3 + 2R_1 R_4 + R_5 \tag{3.9}$$

The length r of the contagiousness period can be estimated from the ARIMA series of the stationary random variables Y_j's, equal to the X_j's without their trend, by considering the length of the interval on which the auto-correlation function remains more than a certain threshold, e.g., 0.1 [101]. For example, by assuming $r = 3$, if $R_1 = a, R_2 = b$ and $R_3 = c$, we obtain:

$$
\begin{aligned}
X_0 &= 1,\ X_1 = a,\ X_2 = a^2 + b,\ X_3 = a^3 + 2ab + c,\ X_4 = a^4 + 3a^2b + b^2 + 2ac, \\
X_5 &= a^5 + 4a^3b + 3ab^2 + 3a^2c + 2bc,\ X_6 = a^6 + 5a^4b + 4a^3c + 6a^2b^2 + 6abc + b^3 + c^2, \\
X_7 &= a^7 + 6a^5b + 5a^4c + 10a^3b^2 + 12a^2bc + 4ab^3 + 3b^2c + 3ac^2 \qquad (3.10)
\end{aligned}
$$

If R_1 and R_2, respectively, to a and b, and if a = b = $R/2$, c = 0, then X_5 is given by:

$$X_5 = \frac{1}{32}R^5 + \frac{1}{4}R^4 + \frac{3}{8}R^3 \qquad (3.11)$$

If $R = 2, \{X_j\}_{i=1,\ldots,\infty}$ is exactly the Fibonacci sequence, and more generally, for $R > 0$, the generalized Fibonacci sequence.
In the general case, where different clusters appear independently at different times, the problem is much more difficult. It can be simplified if it is assumed that these different clusters appear at the same time with the same age of infection. If we suppose that there is more than a unique infected, i.e., if the initial number of infected is $I_0 = X_0 > 1$ at time 0 and if these I_0 first infected are infectious at same age of infection at time 0, then the Equation (3.5) can be written as:

$$X_j = \sum_{d=1,\ldots,t} R_d X_{j-d},\ \forall\ j > 0$$

or:

$$X_j = \sum_{d=1,\ldots,r} R_d X_{j-d} \text{ for all } t > 0 \text{ with } X_{j-d} = 0, \text{ if } j - d < 0$$

This problem of the initial conditions is crucial. It is relatively simple in the case of a single infected appearing at day 0 (unique cluster) or in the case of multiple infected individuals but with the same age of infection, in which case the infection dynamics proceeds according to Equation (3.5), where X_j is the number of new infected individuals at day j and R_d is the daily basic reproduction number at the day d of the period of contagiousness of length r (R_d and r being supposed to be the same for all individuals and constant during the growth phase of new cases).
Another problem is the presence of noise on the registered data. Let us consider X_0, the size of the first cluster of infectious individuals as a Gaussian random variable with a normal distribution $N(\mu_0, \sigma_0)$. Then, if R_d represents the daily basic reproduction number at day d of the contagiousness period of length r and if R_d and r, supposed to be not random real numbers are identical for all infectious individuals and constant during the growth epidemic period, we have, for $m = 6$:

$$
\begin{aligned}
X_1 &= R_1X_0 \\
X_2 &= R_1X_1 + R_2X_0 = (R_1^2 + R_2)X_0 \\
X_3 &= R_1X_2 + R_2X_1 + R_3X_0 = (R_1^3 + 2R_1R_2 + R_3)X_0 \\
X_4 &= R_1X_3 + R_2X_2 + R_3X_1 + R_4X_0 = (R_1^4 + 3R_1^2R_2 + 2R_1R_3 + R_2^2 + R_4)X_0 \\
X_5 &= R_1X_4 + R_2X_3 + R_3X_2 + R_4X_1 + R_5X_0 \\
&= (R_1^5 + 4R_1^3R_2 + 3R_1^2R_3 + 3R_1R_2^2 + 2R_2R_3 + 2R_1R_4 + R_5)X_0 \\
X_6 &= R_1X_5 + R_2X_4 + R_3X_3 + R_4X_2 + R_5X_1 + R_6X_0 \\
&= (R_1^6 + 5R_1^4R_2 + 4R_1^3R_3 + 3R_1^2R_4 + 6R_1^2R_2^2 + 6R_1R_2R_3 + R_2^3 + R_3^2 + 2R_2R_4 + 2R_1R_5 + R_6)X_0 \\
X_7 &= R_1X_6 + R_2X_5 + R_3X_4 + R_4X_3 + R_5X_2 + R_6X_1
\end{aligned}
$$

and more generally:

$$X_t = \sum_{d=1,\ldots,r} R_d X_{t-d} \text{ for } \ t \geq 0 \text{ with } X_{d-k} = 0, \text{ for } d < k, \text{ then } X_t = K(t)X_0$$

and

$$\mu_t = \sum_{d=1,\ldots,r} R_d \ \mu_{t-d}, \text{ for } \ t \geq 0 \text{ with } \mu_{d-k} = 0, \text{ for } d < k, \text{ then } \mu_t = K(t)\mu_0$$

where μ_t is the expectation of X_t and $K(t)$ the general term of the generalized Fibonacci (or discrete Volterra) series. Because X_0 is Gaussian, all the X_t's are Gaussian with distribution $N(K(t)\mu_0, K(t)\sigma_0)$. If we consider now the vectors $\mathbf{R} = (R_1, \ldots, R_r)^T, \mathbf{X_t} = (X_t, \ldots, X_{t-r+1})^T$ and $\mu_\mathbf{t} = (\mu_t, \ldots, \mu_{t-r+1})^T$ for $t > r$, then we have:

$$\mathbf{X_t} = M_t \mathbf{R}$$

where the matrix M_t is given by:

$$M_t = \begin{bmatrix} X_{t-1} & X_{t-2} & \cdots & X_{t-r} \\ X_{t-k-1} & X_{t-k-2} & \cdots & X_{t-k-r} \\ & & \cdots & \\ X_{t-r} & X_{t-r-1} & \cdots & X_{t-2r+1} \end{bmatrix}$$

and, for the expectations:

$$\mu_\mathbf{t} = \mathbf{M_t} \mathbf{R}$$

where the matrix $\mathbf{M_t}$ is given by:

$$\mathbf{M_t} = \begin{bmatrix} \mu_{t-1} & \mu_{t-2} & \cdots & \mu_{t-r} \\ \mu_{t-k-1} & \mu_{t-k-2} & \cdots & \mu_{t-k-r} \\ & & \cdots & \\ \mu_{t-r} & \mu_{t-r-1} & \cdots & \mu_{t-2r+1} \end{bmatrix}$$

If random variable X_t's are observed, a deconvolution of equations above allows obtaining the vector R.
The deconvolution of the first matrix equation leads to random values of the daily basic reproduction numbers R_d and of the second leads to deterministic values of the numbers R_d, but it is clear that the first matrix equation can only be obtained by simulation or by observing epidemic in different countries supposed to have the same spread characteristics (concerning in particular r and R_d parameters), which requires extensive investigation work on reliable data all around the world.
As regards the case of several clusters at the start of the epidemy with different ages of infection, a model generalized to multiple cohorts would make it possible to consider an estimation of the R_d's for the various initial clusters, if it would be possible to define a non-degenerate deconvolution operator.

3.2.6 A stochastic contagion discrete mechanism: obtained from the SIR equation

In the following, we suppose that the susceptible population size remains constant, which constitutes a hypothesis valid during the exponential phase of epidemic waves. The Markovian stochastic approach is linked by a common background consisting of the birth and death

process approach used in the kinetics of molecular reactions by Delbrück [105], then in dynamical systems theory by numerous authors [106, 107], namely in modeling of the epidemic spread in exponential growth. In the Markovian approach, the Malthusian parameter is equivalent to the Kolmogorov–Sinai entropy (called evolutionary entropy in [76, 77, 120]). Another way to derive the SIR equation is the probabilistic approach, which comes from the microscopic equation of molecular shocks by Delbrück [105] and corresponds to a classical birth-and-death process: if at least one event (with rates of contact ν, birth f, death μ or recovering ρ) occurs in the interval $(t, t+dt)$, and by supposing that births compensate deaths, leaving constant the total size N of the population, we have:

$$\begin{aligned} & P(\{S(t+dt)=k, I(t+dt)=N-k\}) \\ &= P(S(t)=k, I(t)=N-k)[1-[\mu k+\nu k(N-k)-fk-\rho(N-k)]dt] \\ &\quad +P(S(t)=k-1, I(t)=N-k+1)[f(k-1)+\rho(N-k+1)]dt \\ &\quad -P(S(t)=k+1, I(t)=N-k-1)[\mu(k+1)+\nu(k+1)(N-k-1)]dt \end{aligned} \tag{3.12}$$

Hence, we have, if $P_k(t)$ denotes Probability$(\{S(t)=k, I(t)=N-k\})$:

$$\begin{aligned} \frac{dP_k(t)}{d} &= [P(S(t+dt)=k, I(t+dt)=N-k)-P(S(t)=k, I(t)=N-k)]/dt \\ &= -P(S(t)=k, I(t)=N-k)[\mu k+\nu k(N-k)n-fk-\rho(N-k)] \\ &\quad +P(S(t)=k-1, I(t)=N-k+1)[f(k-1)+\rho(N-k+1)] \\ &\quad -P(S(t)=k+1, I(t)=N-k-1)[\mu(k+1)+\nu(k+1)(N-k-1)], \end{aligned}$$

and we obtain:

$$\begin{aligned} \frac{dP_k(t)}{dt} &= -[\mu k+\nu k(N-k)-fk-\rho(N-k)]P_k(t)+[f(k-1)+\rho(N-k+1)]P_{k-1}(t) \\ &\quad -[\mu(k+1)+\nu(k+1)(N-k-1)]P_{k+1}(t) \end{aligned}$$

Then, by multiplying by s_k and summing over k, we obtain the characteristic function of the random variable S. If births do not compensate deaths, we have:

$$\begin{aligned} & P(\{S(t+dt)=k, I(t+dt)=j\}) \\ &= P(S(t)=k, I(t)=j)(1-[\mu k+\nu kj-fk-\rho j]dt) \\ &\quad +P(S(t)=k-1, I(t)=j+1)[f(k-1)+\rho(j+1)]dt \\ &\quad -P(S(t)=k+1, I(t)=j-1)[\mu(k+1)+\nu(k+1)(j-1)]dt \end{aligned}$$

If S and I are supposed to be independent and if the coefficients ν, f, μ and ρ are sufficiently small, S and I are Poisson random variables [121], whose expectations $E(S)$ and $E(I)$ verify:

$$\frac{dE(S)}{dt} = fE(S)-\nu E(SI)-\mu E(S)+\rho E(I),$$

or, if $f=\mu$,

$$\frac{dE(S)}{dt} \approx E(I)[-\nu E(S)+\rho],$$

leading to the SIR equation for the variables S, I and R considered as deterministic:

$$\begin{aligned} \frac{dS}{dt} &= -\nu SI+\rho R, \\ \frac{dI}{dt} &= \nu SI-kI-\mu I, \\ \frac{dR}{dt} &= kI-\rho R \end{aligned} \tag{3.13}$$

3.2.7 A stochastic contagion discrete mechanism: obtained from Boolean networks

If we choose now a stochastic Hopfield-like transition for the calculation of the Boolean weight state $O_i(t+1)$ of an individual i (equal to 0 in case of normal weight and to 1 in case of overweight or obesity) as a function of $O_i(t)$, by using a potential function P_i:

$$P_i(t) = \sum_{k \in N(i)} W_{ik} \frac{O_k(t)}{T_k} \tag{3.14}$$

where $N(i)$ is the set of nodes linked to i in the interaction network between the individuals of the studied population, T_i is a tolerance parameter (varying between 0 and $+\infty$) quantifying the level of indifference shown by i to the weight state of the nodes belonging to his neighbourhood $N(i)$, and W_{ik} is the interaction coefficient measuring the influence (positive or negative) individual k has on individual i (its value can be equal to 1, 0 or -1, in absence of precise estimation of this influence). Then, the stochastic transition rule with H the Heaviside function ($H(x) = 1, \text{ if } x \geq 0; H(x) = 0, \text{ if } x < 0$) is as follows:

$$\begin{cases} O_i(t+1) = H(P_i(t)), & \text{if } T_i = 0 \\ Probability(\{O_i(t+1) = 1\}) = \dfrac{e^{P_i(t)}}{1 + e^{P_i(t)}}, & \text{if } T_i > 0 \end{cases} \tag{3.15}$$

In Equation (3.15), the function P_i is the analogue of the Hamiltonian function in the Hopfield model [114], and the tolerance parameter T_i is analogue to the temperature: higher the indifference, nearer 1/2 the probability to become obese, which corresponds to a quasi-absence of influence on the individual i by his neighbours regardless of their weight status.
There are four classical types of centrality in an interaction graph G with n nodes (like the graph of Figure 3.1 Top right). The first is the betweenness centrality, defined for a node k as follows [115]:

$$C_B(k) = \sum_{i \neq j \neq k \in G} \frac{\beta_{ij}(k)}{\beta_k} \tag{3.16}$$

where $\beta_{ij}(k)$ is the total number of shortest paths from node i to node j that pass-through node k, and $\beta_k = \sum_{i \neq j \in G} \beta_{ij}(k)$.
The second type of centrality is the degree centrality, defined from the notions of out-, in- or total-degree of a node i, corresponding to the number of arrows of the interaction graph, respectively outing from or entering in the node i, or both. For example, the total-degree centrality is defined by the following formula:

$$C_i^{in-deg} = \frac{\sum_{j=1,\ldots,n; j \neq i} |a_{ij}|}{n-1} \tag{3.17}$$

where $a_{ij} = sign(W_{ij})$ denotes the general coefficient of the signed incidence matrix A of the graph G.
The third type of centrality is the closeness centrality. The closeness is the inverse of the average farness, which is defined by averaging over all nodes j of the network else than i the distance between i and $j \neq i$:

$$C_i^{clo} = \frac{n-1}{\sum_{j=1,\ldots,n; j \neq i} L(i,j)} \tag{3.18}$$

where the distance chosen here is $L(i,j)$, the length of the closest path between i and j.
The last type of classical centrality is the spectral centrality or eigen-centrality, which takes into account the fact that neighbours of a node i can be also highly connected to the rest of the graph, considering that connections to highly connected nodes contribute to the centrality of i more than connections to weakly connected nodes. Hence, the eigenvector centrality of the node i measures more the global influence of i on the whole network and verifies [122]:

$$C_i^{eigen} = \frac{\sum_{j \in N(i)} W_{ij} V_j}{\lambda} \tag{3.19}$$

where λ is the greatest eigenvalue of the incidence matrix of the graph and V its eigenvector.
The four centralities above can be very different (Figure 3.8), but they have each their intrinsic interest: i) betweenness centrality relates to the global connectivity with all nodes of the network, ii) degree centralities corresponds to the local connectivity with only nearest nodes, iii) closeness centrality measures the relative proximity with other nodes for a given distance on the interaction graph, and iv) spectral centrality corresponds to the ability to be connected to possibly a few number of nodes, but having a high connectivity, for example important hub-relays controlling wide sub-networks. Despite the already complementary properties of the classical centralities, we introduce a new notion of centrality, called entropy centrality, taking into account the heterogeneity of the distribution of states of the neighbours of a node i, and not only its connectivity in the graph, and highlighting the symmetric sub-graphs:

$$C_i^{entropy} = -\sum_{k=1,\ldots,s_i} \nu_k \log_2 \nu_k \tag{3.20}$$

where ν_k denotes the k^{th} frequency among s_i frequencies of the histogram of the observed state values in the neighbourhood $N(i)$ of the node i, set of the nodes out- or in-linked to the node i.

3.2.8 A stochastic contagion discrete mechanism: application to co-morbidity risk factor (obesity)

We are interested in modeling the social contagion mechanism through which the obesity can propagate from individual to individual in same or near classes from a high school through the pupil population, pupils changing their weight under the influence of their school friends connected to them through a declared friendship link (possibly not reciprocal).

Obesity in the friendship graph of a Tunisian high school

In Figure 3.8 , we have represented on the left the observed friendship interactions graph in two classes (of level 4 and 5) of a Tunisian high school with 274 pupils having 524 friendship links between them [123, 124, 125]. The Tunisian sample contains 18 individuals overweight or obese. We will represent first the friendship network (Figure 3.8a) by putting in blue the links between the nodes of the network, which represent the school children in question.
The links are directed, and they can be unique between two nodes, when a friendship without reciprocity has been detected (for the sake of simplicity, arrows have not been represented, but in a couple of potential friends, the final node has in general an area more important than the first one). To visually represent the connectivity of the network, we have chosen

to give the nodes a surface proportional to their *in-degree* centrality in Figure 3.8a,b, *total-degree* centrality in Figure 3.8c, and *spectral* centrality in Figure 3.8d. In the three last representations, we have leaved the friendship network evolve to its asymptotic state following the Equation (3.15), we have kept the proportionality of the node area to its centrality and put together the nodes of the network having their centrality in the same value interval. As we can see on Figure 3.8b and Figure 3.8c, the *total-degree* and *in-degree* centralities give about the same feature, but for the *spectral* centrality in Figure 3.8d, the neighbors of a node i play contribute more to its area if they are highly connected to the other nodes, because these second level highly connected nodes contribute to the *spectral* centrality of i more than the weakly connected ones.
We see in Figure 3.8 that the spectral centrality (b) gives priority to the nodes of the contagion network (a) connected to hubs on the left bottom part of the contagion network. On the contrary, the total degree centrality (c) favors all the hubs even when they are connected to a relatively small number of nodes and the in-degree centrality (d) gives an importance to all the nodes depending to the number of other nodes expressing their friendship for them.
If the aim of such a study is to diminish the number of obese, the identification of the obese hubs is important: a nutritional reeducation on a part of these obese or overweight hubs can reduce drastically the number of obese or overweight in their neighbourhood. For example, if we diminish their influence by increasing their tolerance parameter T_i, we can suppress this influence, then reduce the risk to become obese at their contact. Suppose that we start with an interaction network of 100 individuals including 100 obese or overweight and zero normal weight, with interaction weights w_{ij} chosen at random either 0 or 1.
The Figure 3.9 shows that, by increasing the value of the tolerance parameter T_i for the 8 most connected nodes, the mean percentage of normal weight individuals increases until the value $1/2$, corresponding to a random choice equilibrated between the normality on one hand, and overweight or obesity on the second hand. That shows that the states of the nodes of the network can be regularized to the normality after a campaign targeted at the most influential obese or overweight individuals of the interaction network.

Obesity in the friendship graph of a French high school

The top image of Figure 3.10 shows the main connected component (89 nodes) of a friendship interaction graph is observed in two classes of a French high school with 104 pupils having 348 friendship links between them [123, 124, 125]. The French sample contains 17 overweight or obese individuals.
We see in Figure 3.10 that the spectral centrality (middle-left) of the French sample friendship interaction graph gives priority to the big hubs and nodes connected to these hubs. On the contrary, the total-degree centrality (middle-right) favors all hubs, even those connected to a relatively small number of nodes.
The entropy centrality identifies nodes having different states in their neighbourhood and gives them an importance higher than for the nodes having only one type of state in their neighbourhood. Once stabilized after changing the state of the 20 most influential nodes (for example after a dedicated nutritional education) on the left and of the 21 most influential in the right image of the bottom of Figure 3.10, we can observe that the stationary state of the network is fully overweight or obese in the first case of 20 state changes, and fully normal in the second case of 21 state changes, showing a transition of the invariant measure of the contagion network.

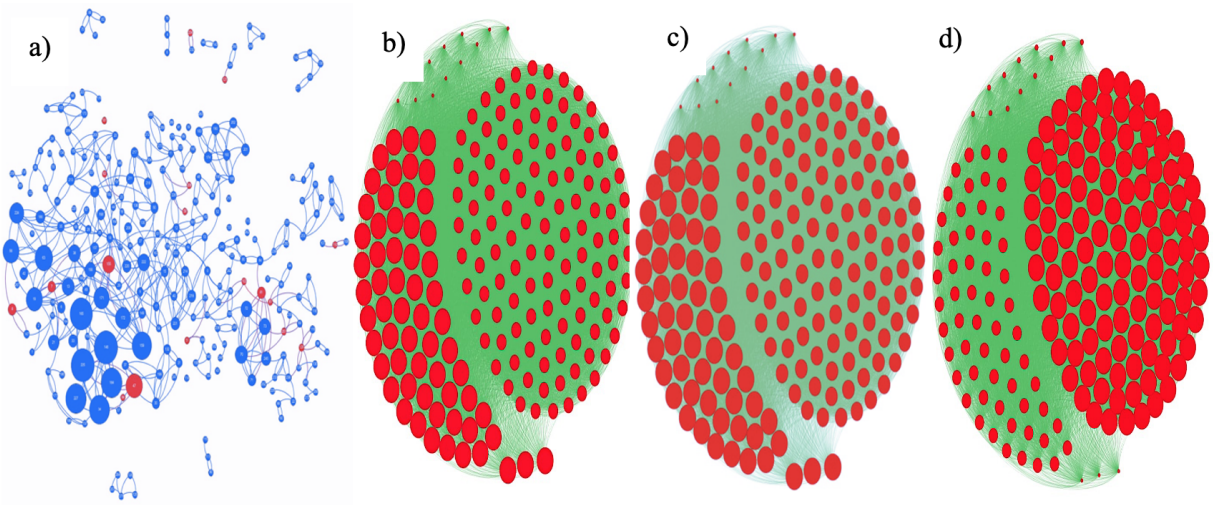


Figure 3.8: a): Friendship empirical network representing a Tunisian sample, with 274 pupils among which 18 overweight or obese (in red) observed from classes of a Tunisian high school, with 524 links between them (only the links having a weight over a fixed threshold have been represented for the sake of visibility). The vertex size is proportional to the *in-degree* centrality of the corresponding node in the friendship interaction graph. b): Representation of the Tunisian friendship interaction graph at its stationary state, iterated from the empirical network as initial state to the asymptotic state under the stochastic transition rule given in Equation (3.15), with node size proportional to its *in-degree* centrality. c): Representation of the *total-degree* centrality of the nodes of the graph b). d): Representation of the *spectral* centrality of the nodes of the graph b).

In the two applications of Boolean dynamics we have presented in this section, we can see some simplifications:

- The small number of states simplifies the dynamics but complicates the examination of realistic solutions to stop the epidemic. We should for example distinguish overweight and obese states in the second case, because the return to a normal weight is much easier for an overweight individual than for an obese one. In the first case, the two-state variable "vaccinated or not" should be added, in order to properly take into account, the effect of the vaccination policy.

- The diffusion through fixed nearest neighbours is often unrealistic, hence it is necessary to build interaction networks with variable links over time, to take into account for example the movement of individuals. This non-constancy of the links can be the subject of a periodic dynamic (linked to school or professional constraints), but it adds a high complexity to the classical Boolean modeling with fixed links over time.

Modeling of the spread of epidemics using Boolean dynamics on an ideal Apollonian network or on a more realistic school network was done by Eric Goles and his colleagues [108]. The current context of comorbidity risk factor (obesity) closely monitored by the WHO, a contagious social pandemic shows that this problem is now highly relevant. The particular chapter of Boolean dynamics applied to pandemics is far from over and the combined efforts of several research communities will continue to make decisive contributions (like we did in this thesis) to this new disciplinary field in the future.

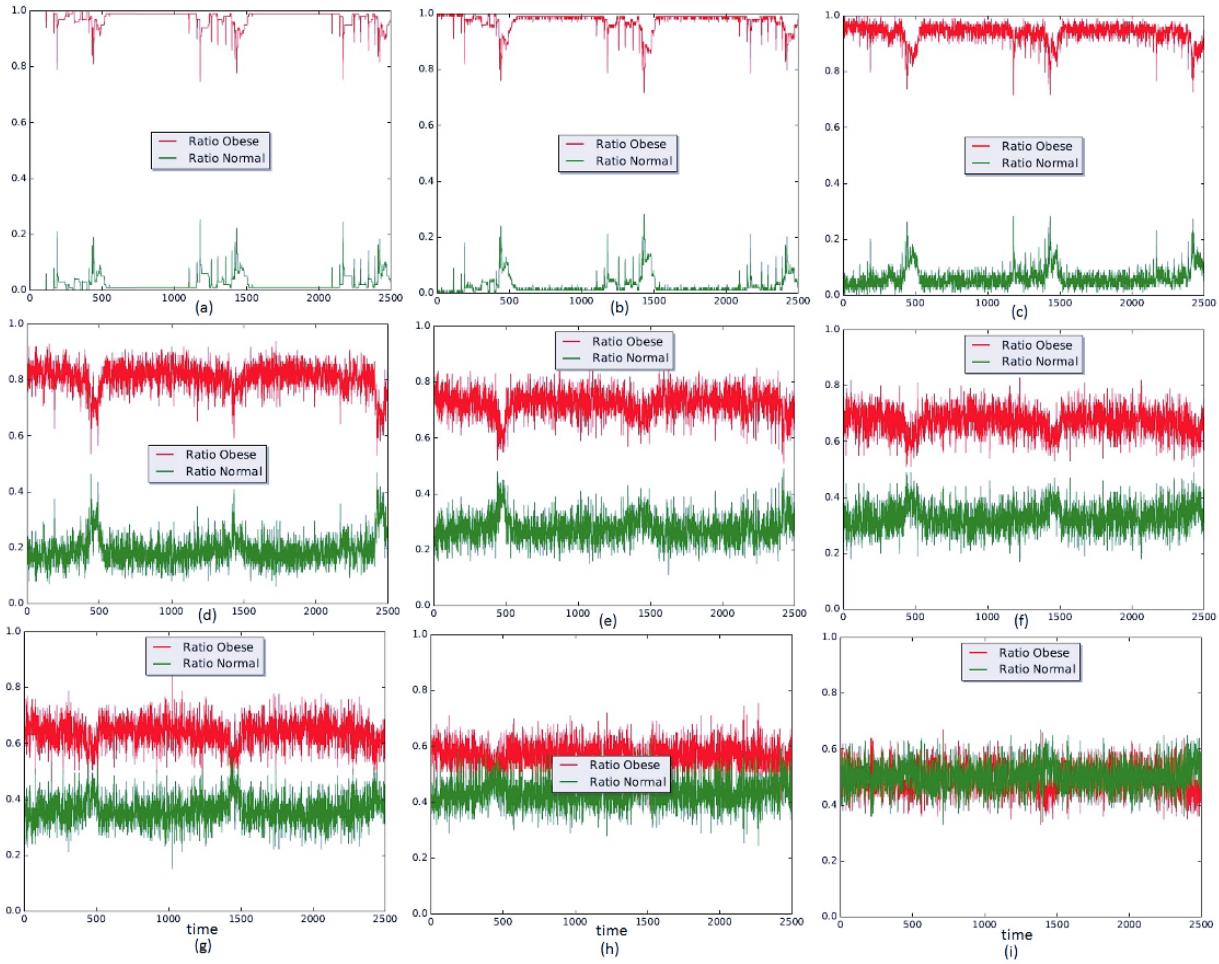


Figure 3.9: Percentages of obese or overweight (red) and normal weight (green) individuals evolving from an initial population with 100 obese or overweight with increasing values of T_i for the 8 individuals with the most neighbours: (a) $T_i = 0.1$, (b) $T_i = 0.5$, (c) $T_i = 1$, (d) $T_i = 2$, (e) $T_i = 3$, (f) $T_i = 4$, (g) $T_i = 5$, (h) $T_i = 10$ and (i) $T_i = 100$, value for which the percentage of normal is $1/2$.

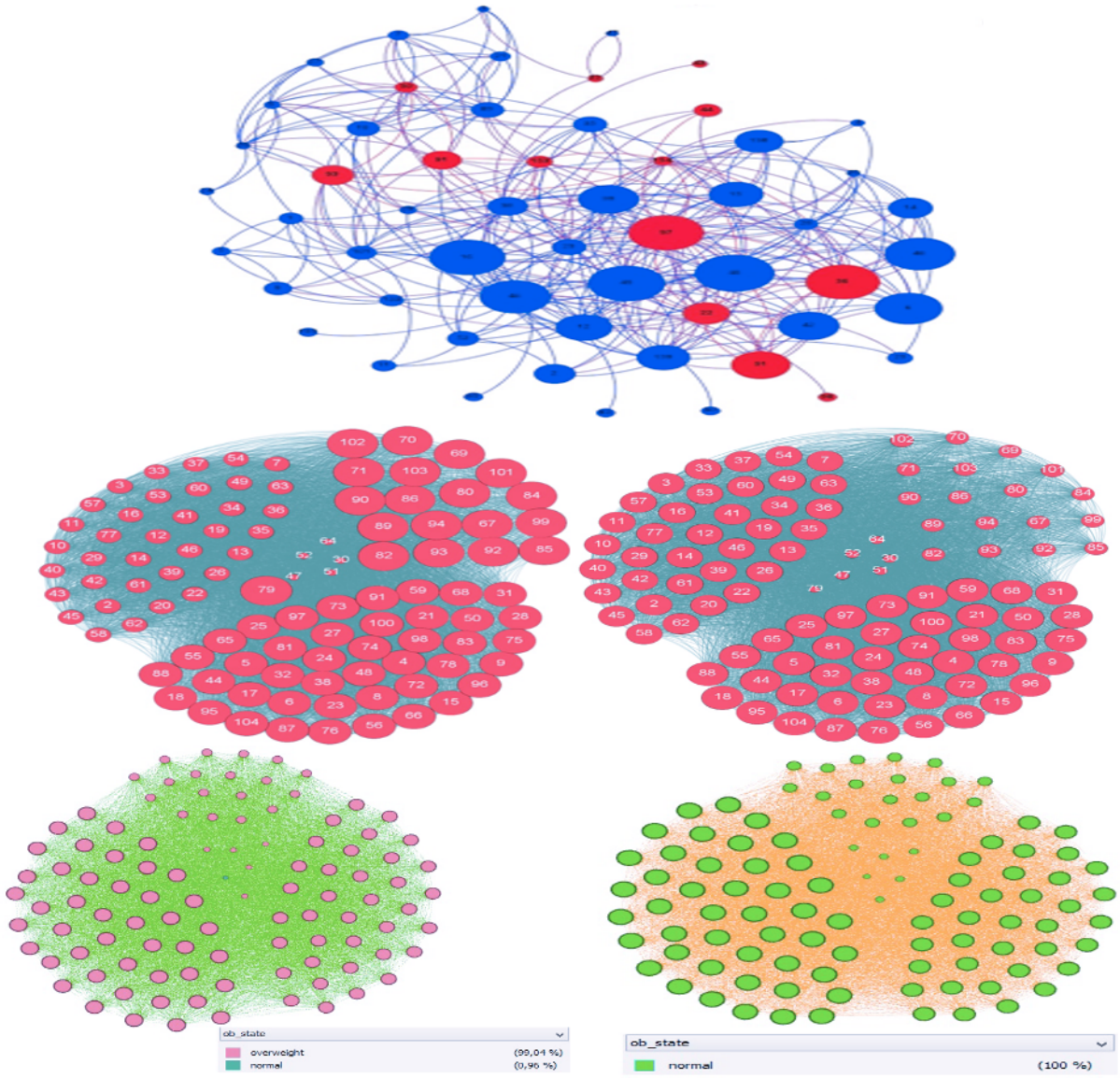


Figure 3.10: **Top**. Friendship empirical network representing the largest connected component of a sample of 104 pupils from which 17 overweight or obese (in red) from a French high school, with 348 links between them (only the links having a weight over a fixed threshold have been represented for the sake of visibility). Node sizes are proportional to their *in-degree* centrality. **Middle**: Friendship interaction graph iterated from the empirical network as initial state to the asymptotic state under the stochastic transition rule given in Equation (3.15), with node size proportional to *spectral* centrality (left) and to *total-degree* centrality (right). **Bottom**: Friendship interaction graph iterated from the empirical graph with *entropy* centrality once stabilized by changing the state of the 20 most entropy-central nodes (left) and of the 21 most entropy-central nodes (right): stationary state is fully overweight or obese in first case and fully normal in the second case, showing a "phase transition" of the asymptotic invariant measure.

3.2.9 Distribution of the daily reproduction numbers along the contagiousness period of an individual: a theoretical approach

If R_0 denotes the basic reproduction number (or average transmission rate) in a given population, we can estimate the distribution V (whose coefficients are denoted $V_j = R_j/R_0$) of the daily reproduction numbers R_j along the contagious period of an individual, by remarking that the number X_j of new infectious cases at day j is equal to $X_j = I_j - I_{j-1}$, where I_j is the cumulated number of infectious at day j, and verifies the convolution equation (equivalent to Equation (3.6):

$$X_j = \sum_{k=1,\ldots,r} R_k X_{j-k}, \text{ giving in continuous time : } X(t) = \int_1^r R(s)\, X(t-s)\, ds, \qquad (3.21)$$

where r is the duration of the contagion period, estimated by $1/(\rho+\mu)$, ρ being the recovering rate and μ the death rate in SIR Equations:

$$\frac{dS}{dt} = -\nu SI, \qquad (3.22)$$

$$\frac{dI}{dt} = \nu SI - (k+\mu)I, \qquad (3.23)$$

where S and I are respectively the size of the susceptible and infectious populations.
Also, r and S can be considered as constant during the exponential phases of the pandemic, and we can assume that the distribution V is also constant; then, V which is an entropy can be estimated as the Kolmogorov-Sinaï entropy of the Markovian Delbrück scheme ruling the X_j's and giving new parameters for characterising pandemic dynamics, namely for quantifying its robustness and stability ([77, 120]) and by solving the linear system (equivalent to Equation (3.7):

$$R = M^{-1}X \qquad (3.24)$$

where M is given by Equation (3.8). Equation (3.24) can be solved numerically, if the pandemic is observed during a time greater than $1/(\rho+\mu)$.
For showing the existence of a unique solution to the deconvolution problem defined to get the daily basic reproduction numbers R_d in simple cases, take the example of $r = 2$, with $R_1 = a$ and $R_2 = b$. In this case, if $X_0 = 1$, the matrix M_5 is defined by:

$$\mathrm{M}_5 = \begin{bmatrix} \mathrm{X}_4 & \mathrm{X}_3 & \mathrm{X}_2 \\ \mathrm{X}_3 & \mathrm{X}_2 & \mathrm{X}_1 \\ \mathrm{X}_2 & \mathrm{X}_1 & \mathrm{X}_\mathrm{o} \end{bmatrix} = \begin{bmatrix} a^4+3a^2b+b^2 & a^3+2ab & a^2+b \\ a^3+2ab & a^2+b & a \\ a^2+b & a & 1 \end{bmatrix}$$

Then, $\det M_5 = a^6 + 3a^4b - a^4 - 2a^3b + 2a^2b^2 - a^2b - 2ab^2$ is a polynomial in a and b, which vanishes on a set of Borel measure zero of $\mathfrak{R}^2$. In the particular case $a = b$, we have $\det M_5 = a^3(a-1)(a^2+4a+3)$ and this set is equal to {0,1}, because the roots of the polynomial a^2+4a+3 are complex conjugates.
The entropy H of V is equal:

$$H = -\sum_{k=1,\ldots,r} V_k \log(V_k)$$

If there are negative V_j's like the Kuwait example in Section 3.2.12, it is still possible to define an index of proximity to uniformity of V by considering the entropy of the distribution W defined by: $W_j = [(V_j - \min\{V_k \leq 0\})/\sum_{i=1,\ldots,r} W_i]$.

We will first demonstrate an example of how the matrix M can be repeatedly calculated for consecutive periods of length equal to that of the contagiousness period (supposed to be constant during the outbreak), giving matrix series $M_1, M_2, \ldots$ Following Equation (3.8), we put the values of X_i's in the two matrices below, with $r = 3$ for two periods, the first from day 1 to day 3 and the second from day 4 to day 6.

$$M_1 = \begin{bmatrix} X_4 & X_3 & X_2 \\ X_3 & X_2 & X_1 \\ X_2 & X_1 & X_0 \end{bmatrix}, \quad M_2 = \begin{bmatrix} X_6 & X_5 & X_4 \\ X_5 & X_4 & X_3 \\ X_4 & X_3 & X_2 \end{bmatrix}, \ldots$$

where, after Equation (3.10), M_1 can be calculated from the R_j's as:

$$M_1 = \begin{bmatrix} R_1^4 + 3R_1^2R_2 + 2R_1R_3 + R_2^2 & R_1^3 + 2R_1R_2 + R_3 & R_1^2 + R_2 \\ R_1^3 + 2R_1R_2 + R_3 & R_1^2 + R_2 & R_1 \\ R_1^2 + R_2 & R_1 & 1 \end{bmatrix},$$

Additionally, from Equation (3.6), if, for instance, $j = 8$ and $r = 3$, then we have the expression below, which means that the new cases on the 8th day depend on the new cases detected on the previous days 7, 6 and 5, supposed to be in a period of contagiousness of 3 days:

$$X_8 = \sum_{k\ =1,2,3} R_k X_{8-k} = R_1X_7 + R_2X_6 + R_3X_5 \tag{3.25}$$

Let us suppose now that the initial R_j's on a contagiousness period of 3 days, are equal to: $\begin{bmatrix} R_1 \\ R_2 \\ R_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$, then matrix M defined by $M_{ij} = X_{7-(i+j)}$ gives the R_j's from Equation (3.24), hence allows the calculation of $X_j = \sum_{k=1,2,3} R_k X_{j-k}$.
The inverse of M is denoted by M^{-1} and verifies: $R = M^{-1}X$, where $X = (X_6, X_5, X_4)$, with $X_1 = 1, X_2 = 2, X_3 = 5, X_4 = 14, X_5 = 37, X_6 = 98$ and we obtain:

$$M_1^{-1} = \begin{bmatrix} 37 & 14 & 5 \\ 14 & 5 & 2 \\ 5 & 2 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} -1/4 & 1 & -3/4 \\ 1 & -3 & 1 \\ -3/4 & 1 & 11/4 \end{bmatrix},$$

and a deconvolution gives the resulting R_j's:

$$\begin{bmatrix} -1/4 & 1 & -3/4 \\ 1 & -3 & 1 \\ -3/4 & 1 & 11/4 \end{bmatrix} \begin{bmatrix} 98 \\ 37 \\ 14 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} R_1 \\ R_2 \\ R_3 \end{bmatrix},$$

thanks to the following calculation:

$$\begin{aligned} R_1 &= -49/2 + 37 - 21/2 = 2 \\ R_2 &= 98 - 111 + 14 = 1 \\ R_3 &= -147/2 + 37 + 77 = 2 \end{aligned}$$

We obtain for the resulting distribution of daily reproduction numbers the exact replica of the initial distribution. We obtain the same result by replacing M_1 by the matrix M_2.

3.2.10 Distribution of the daily reproduction numbers: a stochastic approach

Let us consider a stochastic version of the deterministic toy model corresponding to Equation (3.25), by introducing an increasing noise on the R_j's, e.g., by randomly choosing their values following a uniform distribution on the three intervals: $[2-a, 2+a]$, [1 - a/2, 1 + a/2] and $[2-a, 2+a]$ (for having a U-shape behavior), with increasing values of a, from 0.1 to 1, in order to see when the deconvolution would give negative resulting R_j's, with conservation of the average of their sum R_0, if the random choice of the values of the R_j's at each generation is repeated, following the stochastic version of Equation (3.6):

$$X_j = \sum_{k=1,r} (R_k + \varepsilon_k) X_{j-k},$$

where r is the contagiousness period duration and ε_k is a noise perturbing R_k, whose distribution is chosen uniform on the interval $[0, 2a]$ for $k = 1, 3$, and $[0, a]$ for $k = 2$. This choice is arbitrary, and the main reason of the randomization is to show that the deconvolution can give negative results for R_k's, as those observed for increasing values of a, from 0.1 to 1, with explicit calculations for three consecutive periods, from day 1 to day 3, from day 4 to day 6, and from day 7 to day 9.
For each random choice of the values of the daily reproduction numbers R_j's, we can calculate a matrix M_1 corresponding to Equation (3.7). Its inversion into the matrix M_1^{-1} makes it possible to solve the problem of deconvolution of Equation (3.6)—that is to say, to obtain new R_j's as a function of the observed X_k's. We can then calculate a new matrix M_2 from these new R_j's and thus continue during an epidemic the estimation of the daily reproduction numbers R_j's from the successive matrices $M_1, M_2, \ldots$, and observed X_k's.

1. For $a = 0.1$, let us randomly and uniformly choose the initial distribution of the daily reproduction numbers R_1 in the interval [1.9, 2.1], R_2 in [0.95, 1.05] and R_3 in [1.9, 2.1] as $R_1 = 2.1, R_2 = 0.95, R_3 = 2.1$. Then, the transition matrix M_1 is equal to:

$$M_1 = \begin{bmatrix} 41.7391 & 15.351 & 5.36 \\ 15.351 & 5.36 & 2.1 \\ 5.36 & 2.1 & 1 \end{bmatrix}$$

and we have:

$$M_1^{-1} = \begin{bmatrix} -0.2154195 & 0.92857143 & -0.7953515 \\ 0.92857143 & -2.95 & 1.2178571 \\ -0.7953515 & 1.2178571 & 2.705584 \end{bmatrix}$$

From $X_6 = 113.491, X_5 = 41.7391, X_4 = 15.351$, resulting R_j's are:

$$R_1 = 2.1, R_2 = 0.95 \text{ and } R_3 = 2.1.$$

The next initial R_j's are chosen as: $R_1 = 2, R_2 = 0.95, R_3 = 1.9$ and we have:

$$\begin{aligned} X_7 &= 2X_6 + 0.95X_5 + 1.9X_4 = 226.982 + 39.652 + 29.17 = 295.8 \\ X_8 &= 2X_7 + 0.95X_6 + 1.9X_5 = 591.6 + 107.816 + 79.304 = 778.72 \end{aligned}$$

Then, we obtain the matrices M_2 and M_2^{-1} :

$$M_2 = \begin{bmatrix} 295.8 & 113.491 & 41.7391 \\ 113.491 & 41.7391 & 15.351 \\ 41.7391 & 15.351 & 5.36 \end{bmatrix}$$

$$M_2^{-1} = \begin{bmatrix} -0.07779371 & 0.20964295 & 0.00524305 \\ 0.20964295 & -1.0123552 & 1.26721348 \\ 0.00524305 & 1.26721348 & -3.48354228 \end{bmatrix}$$

Then, the resulting R_j's equal:

$$R_1 = 2.0279, R_2 = 7.6158 \text{ and } R_3 = -16.426.$$

The next initial R_j's are: $R_1 = 2, R_2 = 1.05, R_3 = 1.9$ and we have:

$$\begin{aligned} X_9 &= 2X_8 + 1.05X_7 + 1.9X_6 = 1557.44 + 310.59 + 215.63 = 2083.66 \\ X_{10} &= 2X_9 + 1.05X_8 + 1.9X_7 = 4167.32 + 817.656 + 562.02 = 5546.996 \end{aligned}$$

From these values of X_9 and X_{10}, we obtain the matrices M_3 and M_3^{-1} :

$$M_3 = \begin{bmatrix} 2083.66 & 778.72 & 295.8 \\ 778.72 & 295.8 & 113.491 \\ 295.8 & 113.491 & 41.7391 \end{bmatrix}$$

$$M_3^{-1} = \begin{bmatrix} 0.02596375 & -0.05192766 & -0.04280771 \\ -0.05192766 & 0.0256605 & 0.29823273 \\ -0.04280771 & 0.29823273 & -0.48358035 \end{bmatrix}$$

Then, the resulting R_j's equal:

$$R_1 = 2.486, R_2 = -2.33 \text{ and } R_3 = 7.38769.$$

2. For $a = 1$, let us choose the initial R_1 in [1, 3], R_2 in [0.5, 1.5] and R_3 in [1, 3], e.g., $R_1 = 1, R_2 = 1.355$ and $R_3 = 1.1$. Then, the transition matrix M_1 is equal to:

$$M_1 = \begin{bmatrix} 9.101 & 4.81 & 2.355 \\ 4.81 & 2.355 & 1 \\ 2.355 & 1 & 1 \end{bmatrix}$$

and its inverse is given by:

$$M_1^{-1} = \begin{bmatrix} -1.11983471 & 2.02892562 & 0.60828512 \\ 2.02892562 & -2.93801653 & -1.84010331 \\ 0.60828512 & -1.84010331 & 1.40759184 \end{bmatrix}$$

New cases are: $X_6 = 18.209, X_5 = 9.101, X_4 = 4.81, X_3 = 2.355, X_2 = 1, X_1 = 1$, and by deconvoluting, we obtain the resulting R_j's equal to: $R_1 = 1, R_2 = 1.355, R_3 = 1.1$, i.e., the exact initial distribution.

Let us now consider new initial R_j's: $R_1 = 1, R_2 = 1, R_3 = 1$. That gives a new matrix M_2, with new X_7 and X_8 calculated from the new initial R_j's, by using the former values of $X_6, \ldots, X_2$:

$$\begin{aligned} X_7 &= X_6 + X_5 + X_4 = 18.209 + 9.101 + 4.81 = 32.12 \\ X_8 &= X_7 + X_6 + X_5 = 32.12 + 18.209 + 9.101 = 59.43 \end{aligned}$$

Hence, we obtain:

$$M_2 = \begin{bmatrix} 32.12 & 18.209 & 9.101 \\ 18.209 & 9.101 & 4.81 \\ 9.101 & 4.81 & 2.36 \end{bmatrix}$$

and

$$M_2^{-1} = \begin{bmatrix} -0.35061537 & 0.1839519 & 0.97925345 \\ 0.1839519 & -1.47916605 & 2.31025157 \\ 0.97925345 & 2.31025157 & -8.0783421 \end{bmatrix}$$

and the resulting R_j's equal: $R_1 = 2.90, R_2 = 5.4888, R_3 = -14.696$.

Table 3.1: Simulation results obtained for extreme noises $a = 0.1$ and $a = 1$, showing great variations of deconvoluted distribution of daily reproduction numbers X_j's and a qualitative conservation of their U-shaped distribution along contagiousness period.

a	**Initial R_j's**	**t**	X_t	X_{t+1}	X_{t+2}	**Resulting R_j's**	R_o	**U-shaped**
0.1	2.1;0.95;2.1	4	15.35	31.74	113.5	2.1;0.95;2.1	5.15	Yes
	2;0.95;1.9	6	113.5	295.8	778.7	2.03;7.6;-16.4	-6.77	Inverted
	2;1.06;1.9	8	778.7	2083.7	5547	2.49;-2.33;7.39	7.55	Yes
	1.9;1.05;1.9	10	5547	14207	36776	2.69;-16.7;43.8	29.8	Yes
	1.9;0.95;1.9	12	36776	93910	240359	2.92;1.68;-6.7	-2.1	No
	1.9;1;1.9	14	240359	622149	1605227	2.29;-4.83;14.3	11.8	Yes
	2;1.05;1.9	16	1605227	4331630	11561153	2.76;27;-70	-40.2	Inverted
	1.9;1;1.95	18	11561153	29558395	76502587	2.49;-6.48;17.9	13.9	Yes
	2;1;2.1	20	76502587	207683519	556226772	2.67;-7.6;19.7	14.8	Yes
1	1;1.355;1.1	4	4.81	9.1	18.21	1;1.355;1.1	3.455	Decreased
	1;1;1	6	18.21	32.12	59.43	2.90;5.49;-14.70	-6.31	Inverted
	3;0.5;2.9	8	59.43	247.16	864.34	3.67;-33.9;61.32	31.1	Yes
	2.6;0.7;2.6	10	864.34	2574.82	7942.19	3;-1.79;7.14	8.35	Yes
	2.5;0.75;1.5	12	7942.2	23083.1	67526.6	3.35;2.54;-11.6	-5.71	Decreased
	2.4;0.8;2.4	14	67526.6	199590	588437	2.58;-0.5;4.8	6.88	Yes
	2;1;2	16	588437	1511517	4010652	2.72;-1.08;3.19	4.83	Yes
	2.3;1.15;2.3	18	4010652	12316150	36415885	2.88;-7.9;21.7	16.7	Yes
	2.8;0.6;2	20	36415885	117375471	375133150	3.7;4.1;-17	-9.2	Inverted

We calculate X_9 and X_{10} using new initial R_j's: $R_1 = 3.0, R_2 = 0.5, R_3 = 2.9$:

$$\begin{aligned} X_9 &= 3X_8 + 0.5X_7 + 2.9X_6 = 178.29 + 16.06 + 52.81 = 247.16 \\ X_{10} &= 3X_9 + 0.5X_8 + 2.9X_7 = 741.48 + 29.715 + 93.148 = 864.343 \end{aligned}$$

Hence, we obtain:

$$M_3 = \begin{bmatrix} 247.16 & 59.43 & 32.12 \\ 59.43 & 32.12 & 18.209 \\ 32.12 & 18.209 & 9.101 \end{bmatrix}$$

and

$$M_3^{-1} = \begin{bmatrix} 0.00718287 & -0.00805357 & -0.00923703 \\ -0.00805357 & -0.22288084 & 0.47435642 \\ -0.00923703 & 0.47435642 & -0.80659958 \end{bmatrix}$$

and the resulting R_j's equal:

$$R_1 = 3.66898, R_2 = -33.857 \text{ and } R_3 = 61.32.$$

More precise simulation results are given in Table 3.1, which summarizes computations made for random choices of R_j's distributions, for $a = 0.1$ and $a = 1$ and until time 20. These simulations show a great sensitivity to noise, but a qualitative conservation of their U-shaped distribution along the contagiousness period of individuals. More precisely, because of the presence of noise on the R_j's, we cannot always obtain positive values from the data for the R_j's by applying the deconvolution, which explains the presence of negative values in empirical examples, as in the theoretical noised examples. A way to solve this problem could be to suppose that noise is stationary during all of the growth period of a wave, then calculate the R_j's for all running time windows of length equal to the contagiousness duration and then obtain the mean of the R_j's corresponding to these windows. As this stationary hypothesis is not widely accepted, we prefer to keep negative values and focus on the shape of the distribution of the R_j's.

3.2.11 Distribution of the daily reproduction numbers: application to real data from several countries

A case of 3 days for the duration of the contagiousness period

Figure 3.11 gives the effective transmission rates R_e calculated between 20–25 October 2020 just before the second lockdown in France [126, 127]. As the second wave of the epidemic is still in its exponential phase, it is more convenient (i) to consider the distribution of the marginal daily reproduction numbers and (ii) to calculate its entropy and simulate the epidemic dynamics using a Markovian model [101]. By using the daily new infected cases given in [119], we can calculate, as in Section 3.2.9, the inverse matrix M^{-1} for the period from 20 to 25 October 2020 (exponential phase of the second wave), by choosing 3 days for the duration of contagiousness period and the following raw data for new infected cases: 20,468 for 20 October, then 26,676, 41,622, 42,032, 45,422 and 52,010 for 25 October. Then, for France between 15 February and 27 October 2020, we obtain the daily reproduction numbers given in Figure 3.11 with a U-shape as observed for influenza viruses.

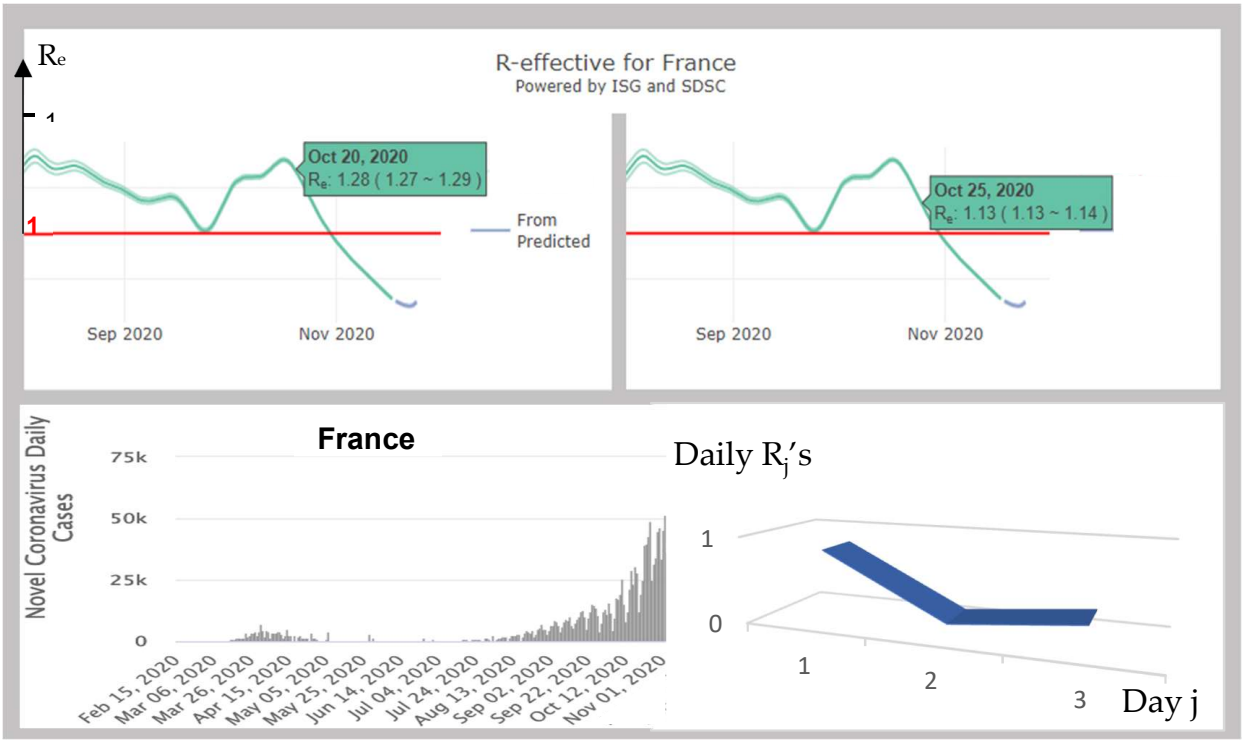


Figure 3.11: **Top**: estimation of the effective reproduction number R_e's for 20 October and the 25 October 2020 (in green, with their 95% confidence interval) [126, 127]. **Bottom left**: daily new cases in France between 15 February and 27 October [119]. **Bottom right**: U-shape of the evolution of the daily R_j's along the 3-day contagiousness period of an individual.

We have:

$$M^{-1} = \begin{bmatrix} 45,422 & 42,032 & 41,622 \\ 42,032 & 41,622 & 26,676 \\ 41,622 & 26,676 & 20,468 \end{bmatrix}^{-1}$$
$$= \begin{bmatrix} -0.0000163989812 & -0.0000292188776 & 0.00007142863 \\ -0.0000292188776 & 0.0000938161392 & -0.0000628537817 \\ 0.00007142863 & -0.0000628537817 & -0.00001447698 \end{bmatrix}$$

Hence, we can deduce the daily R_j's, i.e., the vector (R_1, R_2, R_3) :

$$\begin{bmatrix} -0.0000163989812 & -0.0000292188776 & 0.00007142863 \\ -0.0000292188776 & 0.0000938161392 & -0.0000628537817 \\ 0.00007142863 & -0.0000628537817 & -0.00001447698 \end{bmatrix} \begin{bmatrix} 52,010 \\ 45,422 \\ 42,032 \end{bmatrix}$$
$$= \begin{bmatrix} -0.852911911949567 & -1.32717986039119 & 3.00228812555347 \\ -1.51967382631645 & 4.26131667592337 & -2.64187015405365 \\ 3.71500298367996 & -2.85494447414886 & -0.60849658654673 \end{bmatrix}$$
$$= \begin{bmatrix} 0.82219725466 \\ 0.0997726955533 \\ 0.2515619229844 \end{bmatrix} = \begin{bmatrix} R_1 \\ R_2 \\ R_3 \end{bmatrix}$$

The effective reproduction number is equal to $R_0 \approx 1.174$, a value close to that calculated directly (Figure 3.11), giving $V = (0.7, 0.085, 0.215)$, with a maximal daily reproduction

number the first day of the contagiousness period. The entropy H of V is equal to:

$$H = -\sum_{k=1,\ldots,r} V_k \log(V_k) = 0.25 + 0.21 + 0.33 = 0.79.$$

A case of 6 days for the duration of the contagiousness period

By using the daily new infected cases given in [119] for Chile, we can calculate M^{-1} for the period from 1 to 12 November 2020 (endemic phase), by choosing 6 days for the duration of the contagiousness period and the following 7-day moving average data for the new infected cases (Figure 3.12): 1400 for 1 November, then 1370, 1382, 1359, 1362, 1405, 1389, 1385, 1384, 1387, 1394 and 1408 for 12 November. We have:

$$M^{-1} = \begin{bmatrix} 1394 & 1387 & 1384 & 1385 & 1389 & 1405 \\ 1387 & 1384 & 1385 & 1389 & 1405 & 1362 \\ 1384 & 1385 & 1389 & 1405 & 1362 & 1359 \\ 1385 & 1389 & 1405 & 1362 & 1359 & 1382 \\ 1389 & 1405 & 1362 & 1359 & 1382 & 1370 \\ 1405 & 1362 & 1359 & 1382 & 1370 & 1400 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} -0.05714222 & 0.01016059 & -0.00901664 & 0.01474588 & 0.00640175 & 0.03539322 \\ 0.01016059 & -0.01827291 & 0.0106261 & -0.00763363 & 0.02139586 & -0.01613675 \\ -0.00901664 & 0.0106261 & -0.00544051 & 0.02150289 & -0.01468484 & -0.00286391 \\ 0.01474588 & -0.00763363 & 0.02150289 & -0.01796266 & -0.00553414 & -0.00509801 \\ 0.00640175 & 0.02139586 & -0.01468484 & -0.00553414 & -0.00305831 & -0.00452917 \\ 0.03539322 & -0.01613675 & -0.00286391 & -0.00509801 & -0.00452917 & -0.00686198 \end{bmatrix}$$

Hence, after deconvolution, we obtain:

$$R = \begin{bmatrix} -0.36256122 \\ 0.22645436 \\ 0.01488726 \\ 0.33918287 \\ 0.28557502 \\ 0.50696243 \end{bmatrix}$$

The effective reproduction number is equal to $R_0 \approx 1.011$, a value close to that calculated directly, with a maximal daily reproduction number the last day of the contagiousness period. The quasi-endemic situation in Chile since the end of August, which corresponds to the increase of temperature and drought at this period of the year [101], gives a cyclicity of the new cases occurrence whose period equals the length of the contagiousness period of about 6 days, analogue to the cyclic phenomenon observed in simulated stochastic data of Section 3.2.10. with a similar U-shaped distribution of the R_j's.

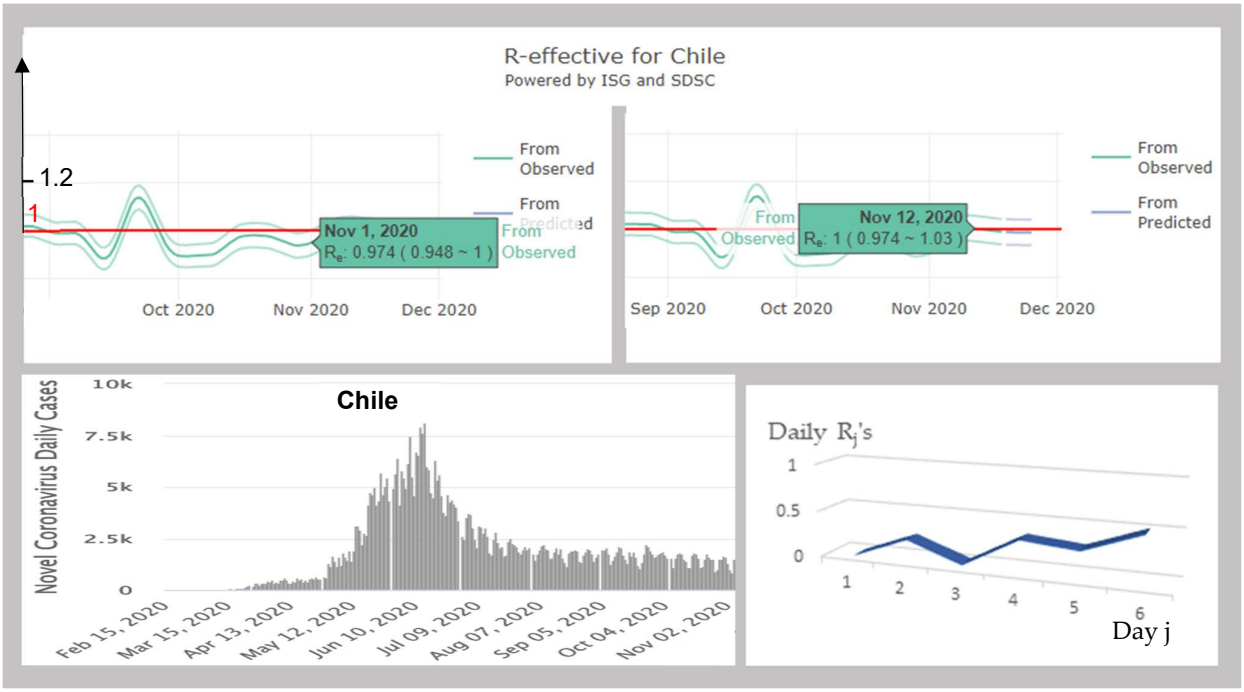


Figure 3.12: **Top**: estimation of the effective reproduction number R_e's for the 1 November and the 12 November 2020 (in green, with their 95% confidence interval) [126, 127]. **Bottom left**: Daily new cases in Chile between 1 November and 12 November [119]. **Bottom right**: U-shape of the evolution of the daily R_j's along the infectious 6-day period of an individual.

A case of 7 days for the duration of the contagiousness period

Let us consider now the beginning of COVID-19 outbreak in USA, for which the estimation of the length of the contagiousness period equals about 7 days ([114, 115, 119]). The numbers of new cases at the start of the disease from February 21 to March 5 2020 are (from [119]): **Feb 21** 20, 0, 0, 18, 4, 3, 0, 3, 5, 7, 25, 24, 34, 63 **March 5**

Then, we have by deconvoluting Equation (3.6):

$$R = \begin{bmatrix} 0.466 \\ 0.584 \\ 1.547 \\ -1.044 \\ 0.174 \\ 0.297 \\ 0.692 \end{bmatrix}$$

The evolution of the R_j's shows on Figure 3.13 a U-shape on day 4 with a sum of the R_j's equal to 2.72, less than the effective reproduction number $R_e = 3.27$ [126]. This U-shape is similar to the classical shape observed for the influenza disease [116].

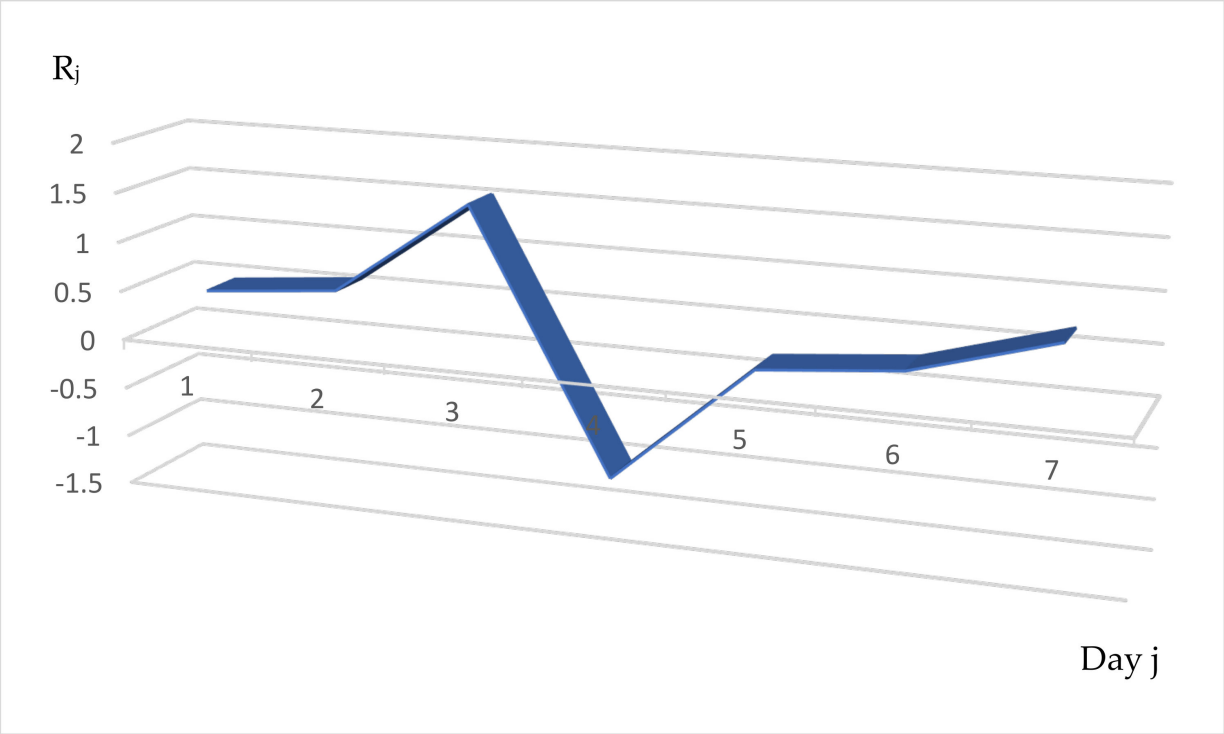


Figure 3.13: Values of the daily reproduction numbers R_j along the period of contagiousness of length 7 days.

3.2.12 The biphasic pattern of distribution of the daily reproduction numbers

A case of U-shaped of the evolution of the daily R_j's

If we use the daily new infectious cases given in [119] during the exponential phase of the second wave of COVID-19 outbreak in Kuwait, we can calculate M^{-1} for the period from December 30 2020 to January 4 2021, by choosing 3 days for the duration of the infectiousness period and the following raw data for the new infected cases are $X_1 = 205$ the 30th December, $X_2 = 286$ the 31th, $X_3 = 285$ the 1st January, $X_4 = 205$ the 2nd, $X_5 = 269$ the 3rd and $X_6 = 372$ the 4th.
Hence, giving for the matrix M :

$$M = \begin{bmatrix} 269 & 205 & 285 \\ 205 & 285 & 286 \\ 285 & 286 & 205 \end{bmatrix}$$

Then, we have:

$$M^{-1} = \begin{bmatrix} 0.00504561 & -0.00852449 & 0.00487808 \\ -0.00852449 & 0.00563046 & 0.00399594 \\ 0.00487808 & 0.00399594 & -0.00747849 \end{bmatrix}$$

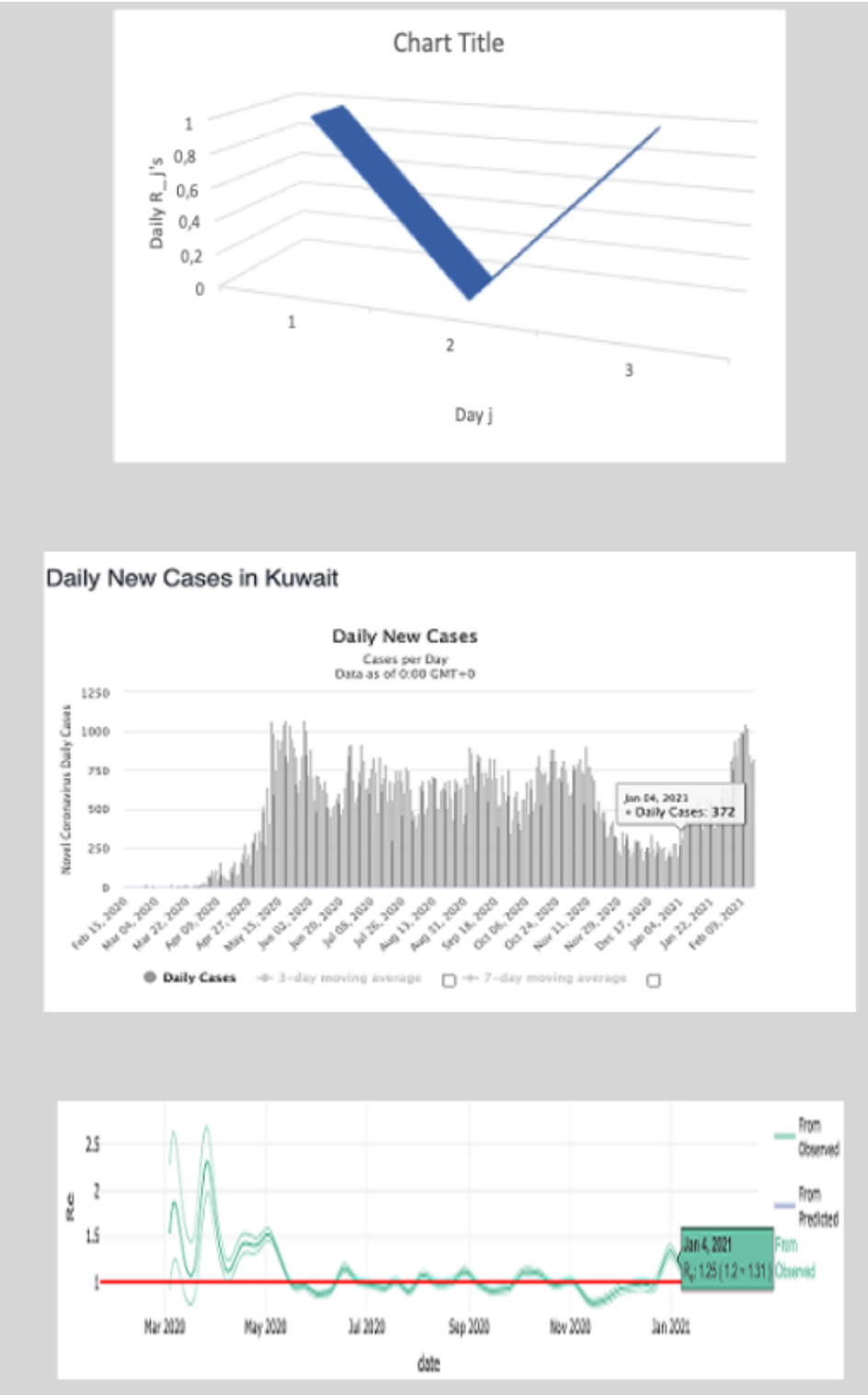


Figure 3.14: U-shape of the evolution of the daily R_j's along the infectious 3-day period of an individual, daily new cases in Kuwait between December 30 2020 and January 4 2021 in worldometer and estimation of the average transmission rate R_o for December 30th 2020 and January 4th 2021 with its 95% confidence interval (in green).

Because

$$X = \begin{bmatrix} X_6 \\ X_5 \\ X_4 \end{bmatrix} = \begin{bmatrix} 372 \\ 269 \\ 205 \end{bmatrix},$$

we get $R = M^{-1}X$, where:

$$\begin{aligned} R_1 &= 0.58388424 \\ R_2 &= -0.83734838 \\ R_3 &= 1.35646161 \end{aligned}$$

Then, we can give a graphical representation of the R_j's (Figure 3.14). The average transmission rate R_o equals about 1.1, value close to that calculated directly, with a maximal daily reproduction rate the first day of the infectiousness period. Because of the negativity of R_2, we have to calculate the entropy H of the distribution

$$W = [(R_1 - R_2)/(R_1 - 2R_2 + R_3), 0, (R_3 - R_2)/(R_1 - 2R_2 + R_3)] = (0.393, 0, 0.607),$$

which is equal to $H = 0.691$, the same value than for the first wave.

A case of increased shape of the evolution of the daily R_j's

By using the daily new infected cases given in [119] for COVID-19 outbreak in Nigeria, we can calculate M^{-1} for the period from 5 November to 10 November 2020 (endemic phase), by choosing 3 days for the duration of the contagiousness period and the following raw data for the new infected cases (Figure 3.15): 141 for 5 November, then 149, 133, 161, 164, and 166 for 10 November.
We have:

$$M^{-1} = \begin{bmatrix} 164 & 161 & 131 \\ 161 & 131 & 149 \\ 131 & 149 & 141 \end{bmatrix}^{-1} = \begin{bmatrix} 0.01796807 & 0.01502897 & -0.03283028 \\ 0.01502897 & -0.02832263 & 0.01575332 \\ -0.03283028 & 0.01575332 & 0.02141264 \end{bmatrix}$$

After deconvolution, we obtain:

$$R = \begin{bmatrix} 0.16177513 \\ 0.38618314 \\ 0.58115333 \end{bmatrix}$$

The effective reproduction number is equal to $R_0 \approx 1.129$, value close to that calculated directly, with a maximal daily reproduction number the last day of the contagiousness period. The distribution V equals (0.143, 0.342, 0.515) and its entropy H is equal to:

$$H = -\sum_{k=1,\ldots,r} V_k \log(V_k) = 0.29 + 0.37 + 0.34 = 1.$$

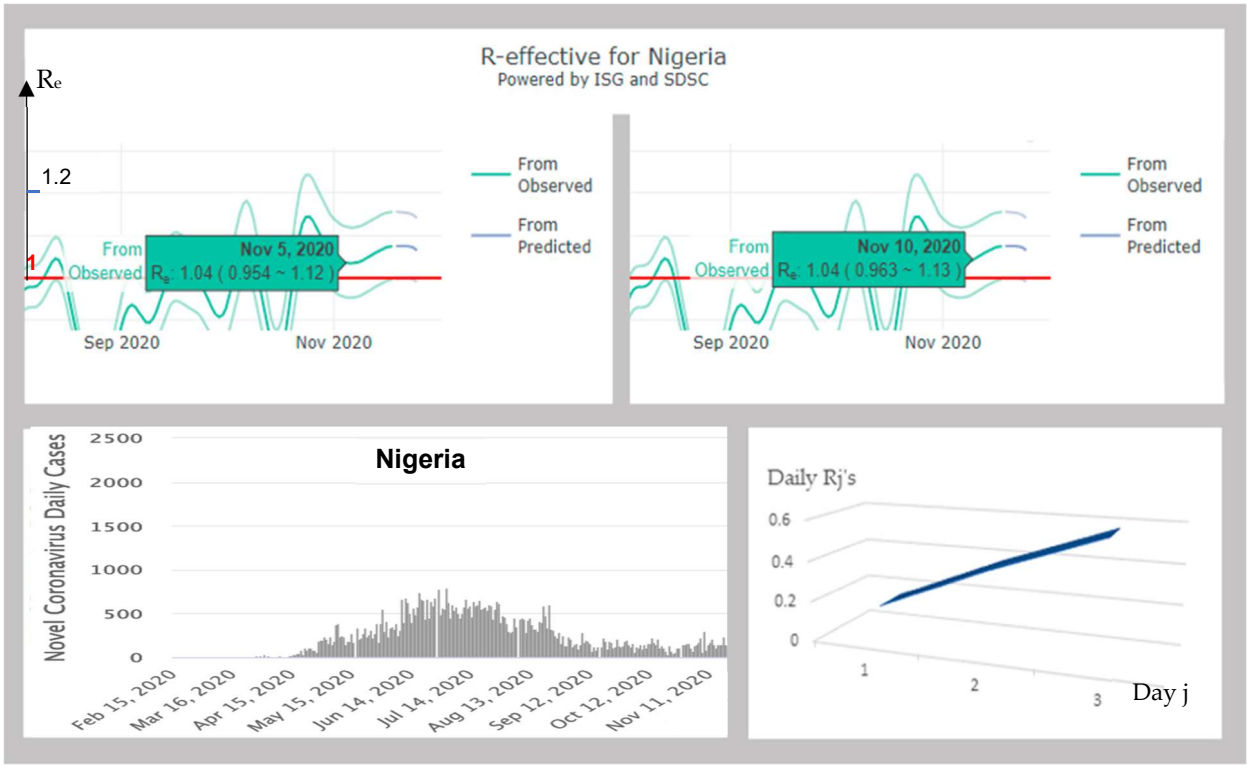


Figure 3.15: **Top**: estimation of the effective reproduction number R_e's for 5 November and 10 November 2020 (in green, with their 95% confidence interval) [126, 127]. **Bottom left**: Daily new cases in Nigeria between 15 February and 21 November [119]. **Bottom right**: increasing evolution of the daily R_j's along the 3-day contagiousness period of an individual.

A case of decreased shape of the evolution of the daily R_j's

By using the daily new infected cases given in [119] for COVID-19 outbreak in Cameroon, we can calculate M^{-1} for the period from 19 June to 24 June 2020 during the first wave, by choosing 3 days for the duration of the contagiousness period and the following raw data for the new infected cases (Figure 3.16): 643 for 19 June, then 329, 281, 149, 229, and 322 for 24 June.

We have:

$$M^{-1} = \begin{bmatrix} 229 & 149 & 281 \\ 149 & 281 & 329 \\ 281 & 329 & 643 \end{bmatrix}^{-1} = \begin{bmatrix} 0.00943779 & -0.00043748 & -0.0039006 \\ -0.00043748 & 0.00889635 & -0.00436075 \\ -0.0039006 & -0.00436075 & 0.00549107 \end{bmatrix}$$

After deconvolution, we obtain:

$$R = \begin{bmatrix} 2.35759568 \\ 1.24664136 \\ -1.43643763 \end{bmatrix}$$

The effective reproduction number is equal to $R_0 \approx 2.17$, value close to effective reproduction number $R_e = 2.56$ [126].

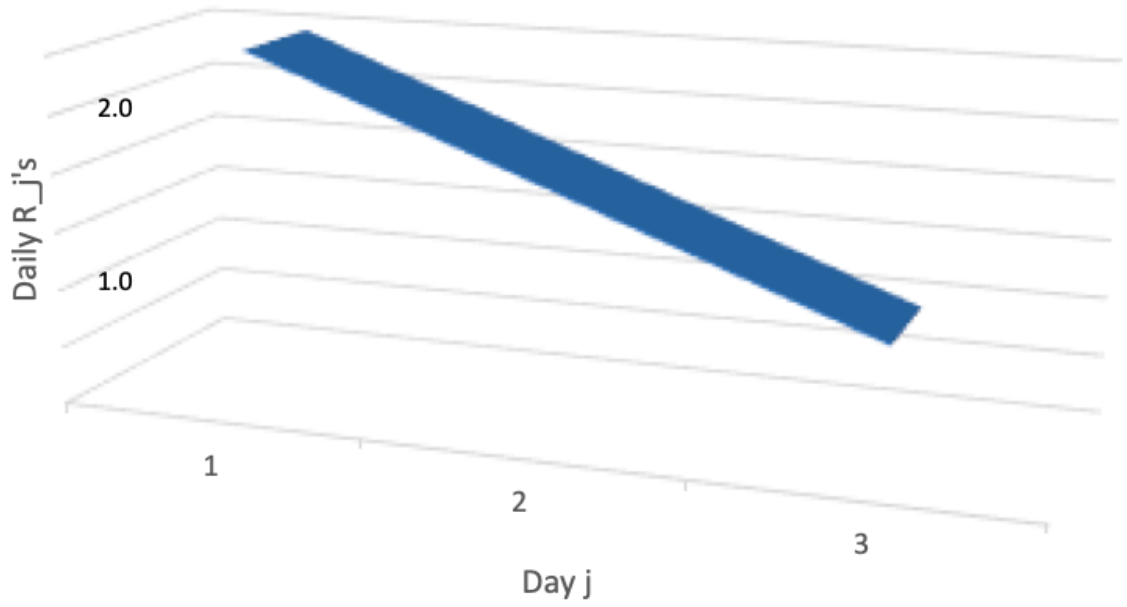


Figure 3.16: Decreasing evolution of the daily R_j's along the infectious 3-day period of an individual, calculated for the period from June 19 to June 24 2020 during the first wave of COVID-19 outbreak in Cameroon.

A case of inverted U-shaped of the evolution of the daily R_j's

By using the daily new infected cases given in [119] for COVID-19 outbreak in Morocco, we can calculate M^{-1} for the period from 20 October to 25 October 2020 during the first exponential phase, by choosing 3 days for the duration of the contagiousness period and the following raw data for the new infected cases (Figure 3.17): 3254 for 20 October, then 3577, 4151, 3685, 4045, and 3020 for 25 October.
We have:

$$M^{-1} = \begin{bmatrix} 4045 & 3685 & 4151 \\ 3685 & 4151 & 3577 \\ 4151 & 3577 & 3254 \end{bmatrix}^{-1} = \begin{bmatrix} -0.00020957 & 0.00084048 & 0.00119126 \\ -0.00084048 & 0.00119679 & -0.00024342 \\ 0.00119126 & -0.00024342 & -0.00094475 \end{bmatrix}$$

After deconvolution, we obtain:

$$R = \begin{bmatrix} 0.35711358 \\ 1.40577244 \\ -0.86841626 \end{bmatrix}$$

The effective reproduction number is equal to $R_0 \approx 0.90$, value close to that calculated directly, with a maximal daily reproduction number the last day of the contagiousness period.

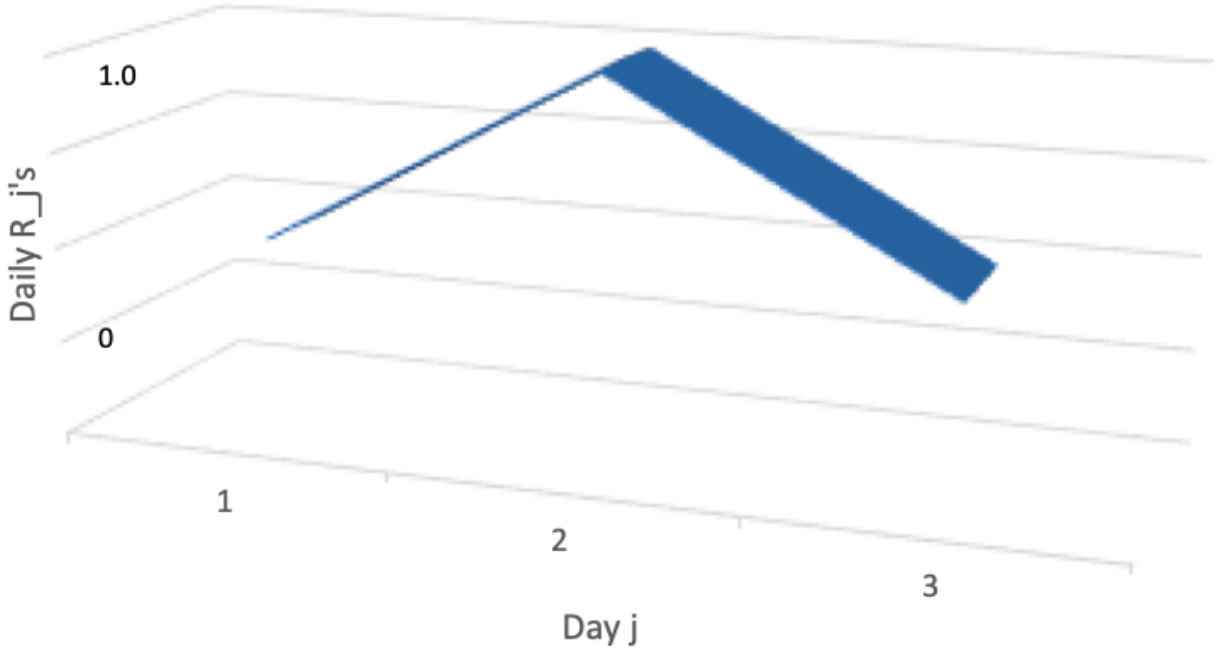


Figure 3.17: Inverted U-shaped evolution of the daily R_j's along the infectious 3-day period of an individual, calculated for the period from October 20 to October 25 2020 during the first exponential phase of COVID-19 outbreak in Morocco.

In Appendix A, Table A.2 gives the shape of the R_j's distribution for 194 countries. Table A.2 is built from new COVID-19 cases at the start of the first and second waves for 194 countries; it shows 42 among these 194 countries having a U-shape evolution of their daily R_j's twice, for 12.12 ± 6 expected with 0.95 confidence ($p < 10^{-12}$), and 189 times, a U-shape evolution for all countries and waves (397), for 99.3 ± 9 expected with 0.95 confidence ($p < 10^{-24}$). Hence, the U-shape is the most frequent evolution of daily R_j's, which confirms the comparison with the behavior of seasonal influenza. The duration of the contagiousness period, as well as the daily virulence, are not constant over time. Three main factors, which are not constant during a pandemic, can explain this:

- In the virus transmitter, the transition between the mechanisms of innate (the first defense barrier) and adaptive (the second barrier) immunity may explain a transient decrease in the emission of the pathogenic agent during the phase of contagiousness [115],
- In the environmental transmission channel, many geophysical factors that vary over time can influence the transmission of the virus (temperature, humidity, altitude, etc.) [101, 102, 103],
- In the recipient of the virus, individual or public policies of prevention, protection, eviction or vaccination, which evolve according to the epidemic severity and the awareness of individuals and socio-political forces, can change the sensitivity of the susceptible individuals [103].

It is therefore very important to seek to estimate the average duration of the period of contagiousness of individuals and the variations, during this phase of contagiousness, of the

associated daily reproduction numbers [122, 128, 129, 130, 131, 132, 133]. If the duration of the contagiousness phase is more than 3–5 days, for example ±7 days, the periodicity of seven days observed for the new daily cases could result of an entrainment of the dynamics of new cases driven by the social "Zeitgeber" represented by the counting of new cases, less precise during the weekend (probably underestimated in many countries not working at this time). That questions the deconvolution over 3 and 5 days, giving some negative R_j. We have extended our study using a duration $r = 3$ of contagiousness to $r = 7$, the result which is the example of USA in Section 3.2.11 show the same existence of identical variations of U-shape type but specify the values of R_j's, more often positive and of more realistic magnitude, while keeping a sum approximately equal to R_0. Rhodes and Demetrius have pointed out the interest of the distribution of the daily reproduction numbers [76] with respect to the classical unique R_0, even time-dependent [77]. In particular, they found that this distribution was generally not uniform, which we have confirmed here by showing many cases where we observe the biphasic form of the virulence already observed in respiratory viruses, such as influenza. The entropy of the distribution makes it possible to evaluate the intensity of its corresponding U-shape. This entropy is high if the daily reproduction numbers are uniform, and it is low if the contagiousness is concentrated over one or two days. If some R_j are negative, it is still possible to calculate this uniformity index of proximity, by shifting their distribution by a translation equal to the inverse of the negative minimum value. We have neglected in the present study the natural birth and death rates by supposing that they are identical, but we could have taken into account the mortality due to the COVID-19. The discrete dynamics of new cases can be considered as Leslie dynamics governed by the matrix equation:

$$X_j = LX_{j-1},$$

where X_j is the vector of the new cases living at day j and L is the Leslie matrix given by:

$$L = \begin{bmatrix} R_1 & R_2 & R_3 & \dots & \dots & R_r \\ b_1 & 0 & 0 & \dots & \dots & 0 \\ 0 & b_2 & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \ddots & \dots & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \dots & \vdots \\ 0 & 0 & 0 & \dots & b_{r-1} & 0 \end{bmatrix} \text{ and } X_{j-1} = \begin{bmatrix} X_{j-1} \\ X_{j-2} \\ X_{j-3} \\ \vdots \\ \vdots \\ X_{j-r} \end{bmatrix},$$

where $b_j = 1 - \mu_j \leq 1, \forall i = 1, \dots, r$, is the recovering probability between days j and $j + 1$. The dynamical stability for L^2 distance to the stationary infection age pyramid

$$P = \frac{\lim_j X_j}{\sum_{i=j,j-r+1} X_i}$$

is related to $|\lambda - \lambda'|$, the modulus of the difference between the dominant and sub-dominant eigenvalues of L, namely $\lambda = e^R$ and λ', where R is the Malthusian growth rate and P is the left eigenvector of L corresponding to λ. The dynamical stability for the distance (or symmetrized divergence) of Kullback–Leibler to P considered as stationary distribution is related to the population entropy H [103, 119, 120, 121, 126, 127], which is defined if $l_j = \prod_{i=1,j-1} b_i$ and $p_j = l_j R_j / \lambda^j$, as follows:

$$H = -\sum_{j=1,\dots,r} \frac{p_j}{\sum_{j=1,\dots,r} j p_j} \log(p_j) \tag{3.26}$$

The mathematical characterization by the population entropy defined in Equation (3.24) of the stochastic stability of the dynamics described by Equation (3.24) has its origin in the theory of large deviations [134, 135, 136]. This notion of stability pertains to the rate at which the system returns to its steady state after a random exogenous and/or endogenous perturbation and it could be useful to quantify further the variations of the distribution of the daily reproduction numbers observed for many countries [137, 138, 139, 140, 141]. In summary, the main limitations of the present study are:

- The hypothesis of spatio-temporal stationarity of the daily reproduction numbers is no longer valid in the case of rapid geo-climatic changes, such as sudden temperature rises, which decrease the virulence of SARS CoV-2 [101], or mutations affecting its transmissibility. The medium of atmospheric, cutaneous or intestinal transmission is of great importance because it depends on numerous environmental conditions, such as humidity.

- The still approximate knowledge of the duration r of the period of contagiousness necessitates a more in-depth study at variable durations, by retaining the value of r, which makes all of the daily reproduction numbers positive.

- The choice of uniform random fluctuations of the daily reproduction numbers is based on arguments of simplicity. A more precise study would undoubtedly lead to a unimodal law varying throughout the contagious period, the average of which following a U-shaped curve, of the type observed in the literature on a few real patients [104, 116, 117]

- Another limitation is the comparison of the results of the simulations with the real behaviour of the epidemic, of which we see, in the COVID-19 outbreak, that most of the parameters of transmissibility and contagiousness can change over time. These changes are linked to [137, 138, 141, 142, 143, 144, 145, 146, 147, 148]:

  i) the source host of the virus (in whom the intensity of the symptoms varies during the period of contagiousness),

  ii) the infectious agent (whose virulence is variable, due for example to the mutations observed),

  iii) the future virus host (in whom, the immunological state depends on multiple factors as age, comorbidities, cross immunity, etc.).

  The long reaction time to mitigation or vaccination measures decided on the basis of the models makes it very difficult to adapt them to the new propagation conditions and considerably limits the operational character of the models, even if their short-term predictions are often of very good quality.

  Among other possible approaches is the use of the Caputo fractional derivative [149], which is particularly interesting. The Caputo fractional derivative of a function f equals

$$D_c^q(f(t)) = \frac{\int_0^T f^n(y)}{(t-y)^{q+1-n}} dy$$

  where the integers n and p verify $n-1 < q < n$.

  The use of the fractional derivative in the models makes it possible to obtain simulations of the increase in the cumulative cases of infected individuals, which fits the data

better than the classical derivative, particularly when the phenomenon of growth of the infected is damped (at the end of an epidemic wave), i.e., after the very early exponential phase where the classical approach presents a slightly lower error than the fractional one [149].

3.3 Maximal Reproduction Number Estimation in COVID-19 Outbreak

3.3.1 Maximal reproduction number model

Suppose that the observation time is continuous and that the new cases of infectious $X(t)$ depend on the distribution of the quantities $O(s) = R(s)/E(R)$, where the quantity $E(R) = \int_{\Re} R(t)dt$:

$$X(t) = \int_{\Re} R(s)X(t-s)ds = E(R)\int_{\Re} X(t-s)O(s)ds \tag{3.27}$$

(a) suppose in a first case that the curve $O(s)$ is a Gaussian curve $\mathcal{N}(m,\sigma)$ then:

$$R(s) = R_{\max}\ \exp\left(-\frac{(s-m)^2}{2\sigma^2}\right), \tag{3.28}$$

where $R_{\max}$ is the maximum of the daily reproduction numbers, and to show that the continuous equivalent of Equation (in Section 3.2) can be solved if $X(t)$ is supposed to be exponential: $X(t) = 100e^{\mathfrak{B}(t-t_0)} = c\,e^{\mathfrak{B}t}$, for all $t \geq t_0$, where $c = 100e^{-\mathfrak{B}t_0}$, Equation (3.28) can be written as:

$$ce^{\mathfrak{B}t} = \int_{\Re} R_{\max}\ \exp\left(-\frac{(s-m)^2}{2\sigma^2}\right)\ c\,e^{\mathfrak{B}(t-s)}ds,$$

from which the following equations can be deduce:

$$1 = R_{\max}\int_{\Re}\exp\left(-\frac{(s-m)^2}{2\sigma^2} - \mathfrak{B}s\right)ds$$

$$1 = R_{\max}\int_{\Re}\exp\left(-\frac{(s-m+\mathfrak{B}\sigma^2)^2}{2\sigma^2} - \mathfrak{B}s\right) + \mathfrak{B}^2\sigma^2/2 - \mathfrak{B}m)\,ds$$

$$1 = R_{\max}\,\sigma(2\pi)^{1/2}\ \exp(\mathfrak{B}^2\sigma^2/2 - \mathfrak{B}m)\int_{\Re}\exp\left(-\frac{(s-m+\mathfrak{B}\sigma^2)^2}{2\sigma^2}\right)/\sigma(2\pi)^{1/2}\,ds$$

Hence,

$$1 = R_{\max}\sigma(2\pi)^{1/2}\ \exp(\mathfrak{B}^2\sigma^2/2 - \mathfrak{B}m) \tag{3.29}$$

and if $\sigma(4\sigma \approx r,$ the duration of the contagiousness period), $m \approx 2\sigma$ and $\mathfrak{B}$(the transmission rate), then $R_{\max}$, the value at day m(the middle of the contagiousness period) of the maximal daily reproduction number.
(b) if the curve of the daily reproduction numbers $R(s)$ is the sum of two Gaussian-shaped curves, the first corresponding to the activation of the innate immunity and the second to the activation of the adaptive immunity, then:

$$R(s) = R_{\max_1}\ \exp(-(s-m_1)^2/2\sigma_1^2 + R_{\max_2}\ \exp(-(s-m_2)^2/2\sigma_2^2$$

in this case, Equation (3.28) can be written as:

$$c\,e^{\mathfrak{B}t} = \left[\int_{\Re} R_{\max_1}\,\exp(-(s-m_1)^2/2\sigma_1^2) + R_{\max_2}\,\exp(-(s-m_2)^2/2\sigma_2^2)\right] c\,e^{\mathfrak{B}(t-s)}\,ds \quad (3.30)$$

Then,

$$1 = R_{\max_1}\,\exp(\mathfrak{B}^2\sigma_1^2/2 - \mathfrak{B}m_1 + R_{\max_2}\,\exp(\mathfrak{B}^2\sigma_2^2/2 - \mathfrak{B}m_2) \quad (3.31)$$

formula from which the calculation of the value of a parameter in function of the value of others. If $4\sigma_1 + 4\sigma_2$ can be estimated by the duration of the contagiousness period and if suppose that σ_1 is close to σ_2, with $m_1 = 2\sigma_1$ and $m_2 = 4\sigma_1 + 2\sigma_2$, then the only remaining unknown parameters are $R_{\max_1}$ and $R_{\max_2}$. If these two maximal daily reproduction numbers are supposed to be close, the Equation (3.31) allows the final calculation of a unique $R_{\max}$, which could be compared to the one obtained by the discrete method of deconvolution presented in Section 3.2.

3.3.2 Application of the model

The method derived in Section 3.3 will first be applied to Cameroon's data concerning the COVID-19 outbreak from the Cameroon Situation Reports (weekly regional new cases and deaths [150]) and from the public database Worldometers (weekly national new cases and deaths [119]) and then to six other countries.

- From Equation (3.29), the maximal daily reproduction number $R_{\max}$ can be estimated

$$1 = R_{\max}\sigma(2\pi)^{1/2}\,\exp(\mathfrak{B}^2\sigma^2/2 - \mathfrak{B}m),$$

Then, from [151], $\sigma = 1, m = 2, \mathfrak{B} = 0.4$. Then,

$$R_{\max} = 1/(2\pi)^{1/2}\,\exp(0.8 - 0.08) \approx 0.9,$$

which corresponds to the case of a Gaussian shape for the daily reproduction numbers $R_j's$ along the contagiousness period of length $r \approx 4\sigma$ to a basic reproduction number

$$R_0 = \sum_{k=1,\ldots,r} R_k \approx 2.1.$$

- To calculate $R_{\max}$ for six other countries using same procedure as above using the transmission rate of the first wave in 2020: Australia $= 4.6214 \times 10^{-8}$, Italy $= 1.2292 \times 10^{-8}$, Spain $= 3.3182 \times 10^{-7}$, United States $= 3.1828 \times 10^{-7}$, United Kingdom $= 1.1526 \times 10^{-7}$ and Canada $= 2.0731 \times 10^{-7}$ given in [152], with $\sigma = 1$ and $\sigma = 2$, corresponding to a duration of the contagiousness period respectively 4 and 8 days. It is also possible to calculate $R_{\max_2}$ by using Equation (3.31) as presented in Table 3.3. The values of $R_{\max}$ in Table 3.2 is used for $R_{\max_1}$ while still maintaining the values for the transmission rate and σ.

Table 3.2: Result for $R_{\max}$ for different countries.

Countries	$\sigma = 1$	$\sigma = 2$
United Kingdom	0.39978	0.20031
Australia	0.40018	0.20071
Italy	0.39927	0.19980
Spain	0.40136	0.20190
Canada	0.40018	0.20098
United States	0.40126	0.20180

Table 3.3: Result for $R_{\max_2}$ while using $R_{\max}$ in Table 3.2 as $R_{\max_1}$ for different countries.

Countries	$\sigma = 1$	$\sigma = 2$
United Kingdom	0.79842	0.59769
Australia	0.79742	0.59608
Italy	0.79970	0.59974
Spain	0.79440	0.59129
Canada	0.79673	0.59498
United States	0.79466	0.59169

3.4 Estimation of Transmission Rate From Identification of the First Inflection Point of New Infectious Cases Waves in COVID-19 Outbreak

Following the same approach as in [153], the inflection point during the outbreak of COVID-19 pandemic can be derived using Bernoulli equation and to analyze this Susceptible–Infected (SI) compartment model for calculating the time of the inflection point of the curve of new infectious cases. First, let's recall that the SI Bernoulli differential equation model of an outbreak epidemic is given as:

$$\begin{aligned} \frac{dS}{dt} &= \frac{\beta}{S_0} S(t)I(t) + dS(t), \\ \frac{dI}{dt} &= \frac{\beta}{S_0} S(t)I(t) - \nu I(t), \end{aligned} \tag{3.32}$$

where d is the susceptible demographic balance, ν the specific mortality rate due to the disease, β/S_0 the disease transmission rate, $S(t)$ the number of susceptible individuals, $I(t)$ the number of infectious individuals at time $t \geq 0$, and the initial conditions of the model are: $S(0) = S_0 > 0$ and $I(0) = 1$.

By simulating Equation (3.32), the results obtained by neglecting the mortality due to the COVID-19 outbreak and to the demography of the susceptible population are given on Figure 3.19.

The Figure 3.18 shows slight differences at the beginning of the epidemic wave between Model 1, corresponding to Equation (3.32) in which the fertility is neglected and the mortality ($d = \nu = 0$) and Model 2, corresponding to Equation (3.32) by fixing the demographic balance of susceptibles d to 0.01 and the specific mortality due to the COVID-19 outbreak ν to 0.01, which are realistic values for Cameroon.

Let's assume we want to simulate Bernoulli equation if it is evolving in time due to various influences as geoclimatic, socio-economic and demographic factors but Equation (3.32) must be rewritten as follows:

$$\begin{aligned} \frac{dS}{dt} &= \mathfrak{B}SI - \mu S + f + \sigma I, \\ \frac{dI}{dt} &= \mathfrak{B}SI - \nu I + f + \rho I, \end{aligned} \tag{3.33}$$

with $S(0) = S_0, I(0) = I_0$ and where $S(t)$ (resp. $I(t)$) is the size of susceptible (resp. infected) at time t, $\mathfrak{B}$ the transmission rate, μ (resp. ν) the natural (resp. viral) death rate,

f the fecundity, σ (resp. ρ) the post-infected non immunized (resp. immunized) rate. If $f = \mu = \rho = \sigma = \nu = 0$, the solution of Equation (3.33) is given by:

$$S(t) = \frac{S_0}{1 + e^{\mathfrak{B}(t+t_0)}} \text{ and } I(t) = \frac{S_0 e^{\mathfrak{B}(t+t_0)}}{1 + e^{\mathfrak{B}(t+t_0)}}, \tag{3.34}$$

with $S + I = S_0$. If $I(0) = 1$, then if $t_0 = -log(S_0 - 1)/\mathfrak{B}, S(0) = S_0/1 + e^{\mathfrak{B}t_0} \approx S_0 - 1$. If S_0 is large, then during the time lapse at which $S(t)$ remains close to S_0, if $\mu S_0 = f$, then Equation (3.33) becomes close to the linear system:

$$\frac{dS}{dt} = (-\mathfrak{B}S_0 + \rho)I \text{ and } \frac{dI}{dt} = (\mathfrak{B}S_0 - \nu - \rho)I, \tag{3.35}$$

with the solution:

$$S(t) = \frac{(-\mathfrak{B}S_0 + \rho)e^{(\mathfrak{B}S_0 - \nu - \rho)t}}{(\mathfrak{B}S_0 - \nu - \rho)} \text{ and } I(t) = e^{(\mathfrak{B}S_0 - \nu - \rho)t}, \tag{3.36}$$

where $\mathfrak{B}S_0$ represents the basic reproduction number of the epidemic disease, usually denoted R_0.

The present models are in general used to simulate an epidemic wave, as an excursion in the phase plane of an ordinary two-dimensional differential system having some excitable ability, i.e., the existence in his phase portrait of large return trajectories after perturbation of its stable stationary state or of its parameters, these trajectories returning to either an endemic or an eradicated final state. The difference between these two asymptotic behaviors is that in endemic case, the final value of the infected, back to the stationary state is non-zero and, in eradicated case, this value is zero. These behaviors are illustrated by simulations of the Bernoulli model in the possible final states corresponding to epidemic (transients in Figure 3.21), eradication (end state in Figure 3.21 top) and endemic (low level end in Figure 3.21 middle and high level in Figure 3.21 bottom).

Suppose that a parameter of the Bernoulli model like the transmission rate $\mathfrak{B}$ is evolving in time due to various influences as geo-climatic factors (temperature, humidity, elevation, etc.), socio-demographic determinant (density and median age of the target population) and economic variables (GDP, Gini's index, inequality index, etc.), the actual curve of the new infected cases observed during the epidemic outbreak is in fact also driven by the slow dynamics of the seasonally varying or linearly growing parameters.

Then, the observed dynamics of new infected cases results from the epidemic dynamics combined with slow evolution of these external factors on variation surfaces mixing both epidemic variables and parameters dynamics. The Figure 3.20 shows such a surface corresponding to the variations of the transmission rate $\mathfrak{B}$ due to climate changes (transition between winter and summer times) or public health policies (quarantine, vaccination, etc.), and the actual (S, I) trajectory lies on the corresponding surface. The influence of these different factors on the variations of the trajectories (S, I) are difficult to take into account in a differential model of the type Equation (3.33), which is why Section 4.5 will treat the statistical approach providing a better understanding of the correlations between the exogenous determinants of the epidemic and the level of new cases of infected.

Graphical representation of the inflection point P_I of the curve of $I(t)$ is given in Figure 3.19, for the first half part of the second wave of the COVID-19 weekly new reported cases between week 5 2021 and week 12 2021 (data from [150]). The red curve represents the cubic spline

approximation of the raw data with the following statistics: p-value = 0.03, Multiple R-squared (MR) = 0.90, Relative Mean Square Error (RMSE) = 0.247.
During the same period, it is also possible to smooth the death data (from [119]) with a cubic spline with the following statistics: p-value = 0.12, Multiple R-squared (MR) = 0.99, Relative Mean Square Error (RMSE) = 4.380 (see Figure 3.22).

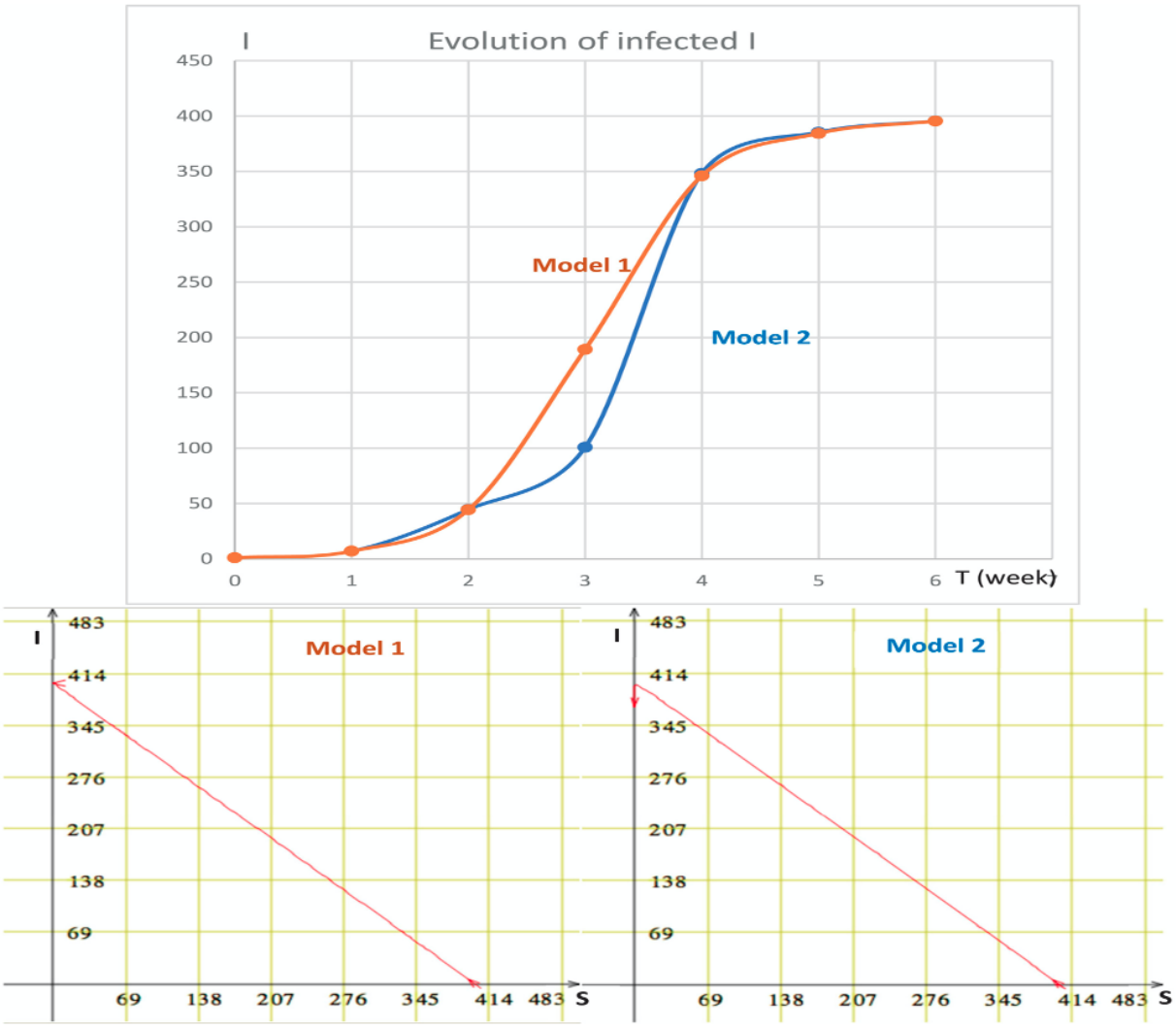


Figure 3.18: Top: simulation of the evolution of the number of new infected I in Model 1 (Equation (3.32) with $d = \nu = 0$) and Model 2 (Equation (3.32) with $d = \nu = 0.01$) Bottom: simulation of Models 1 and 2 with same initial conditions for infected ($I(0) = 1$) and susceptible individuals ($S(0) = 400$).

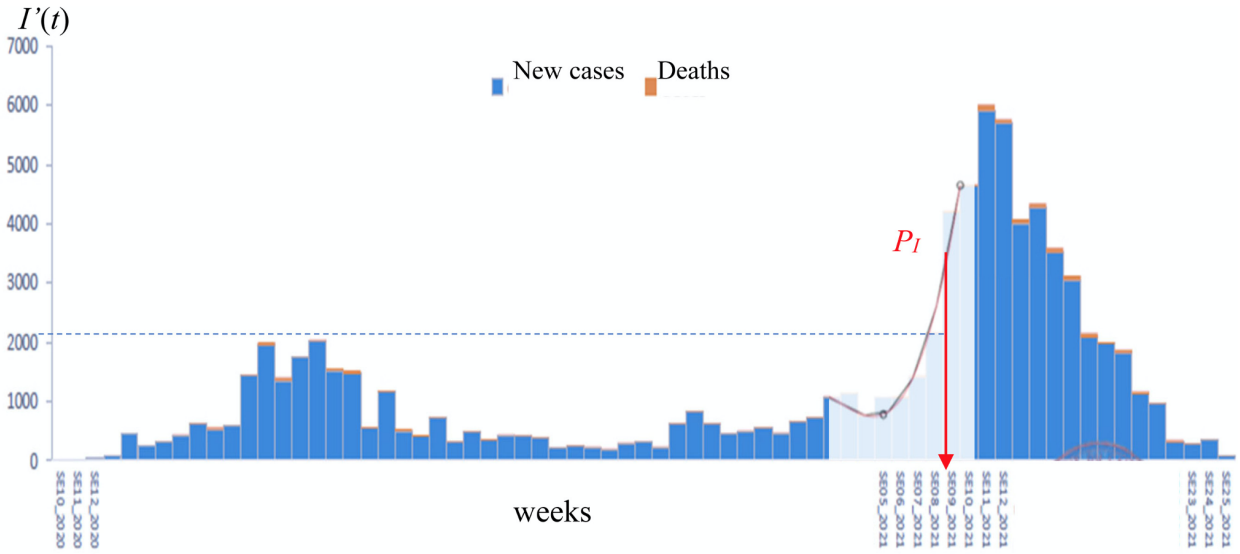


Figure 3.19: COVID-19 weekly new cases in Cameroon showing the inflection point P_I (red arrow) during the first half part of the second wave between week 5 of 2021 and week 12 of 2021. (after [150]).

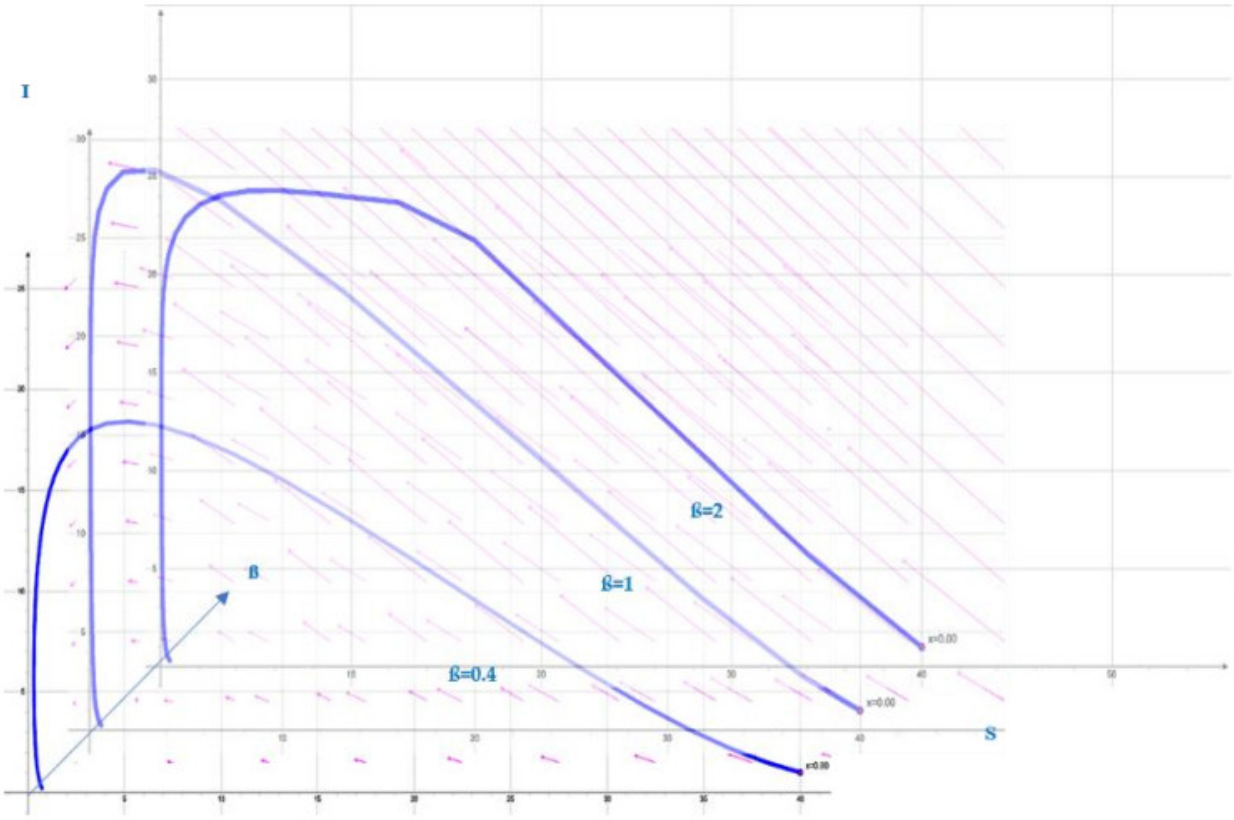


Figure 3.20: Surface of evolution of 3 covariables, S, I and $\mathfrak{B}$, the transmission rate.

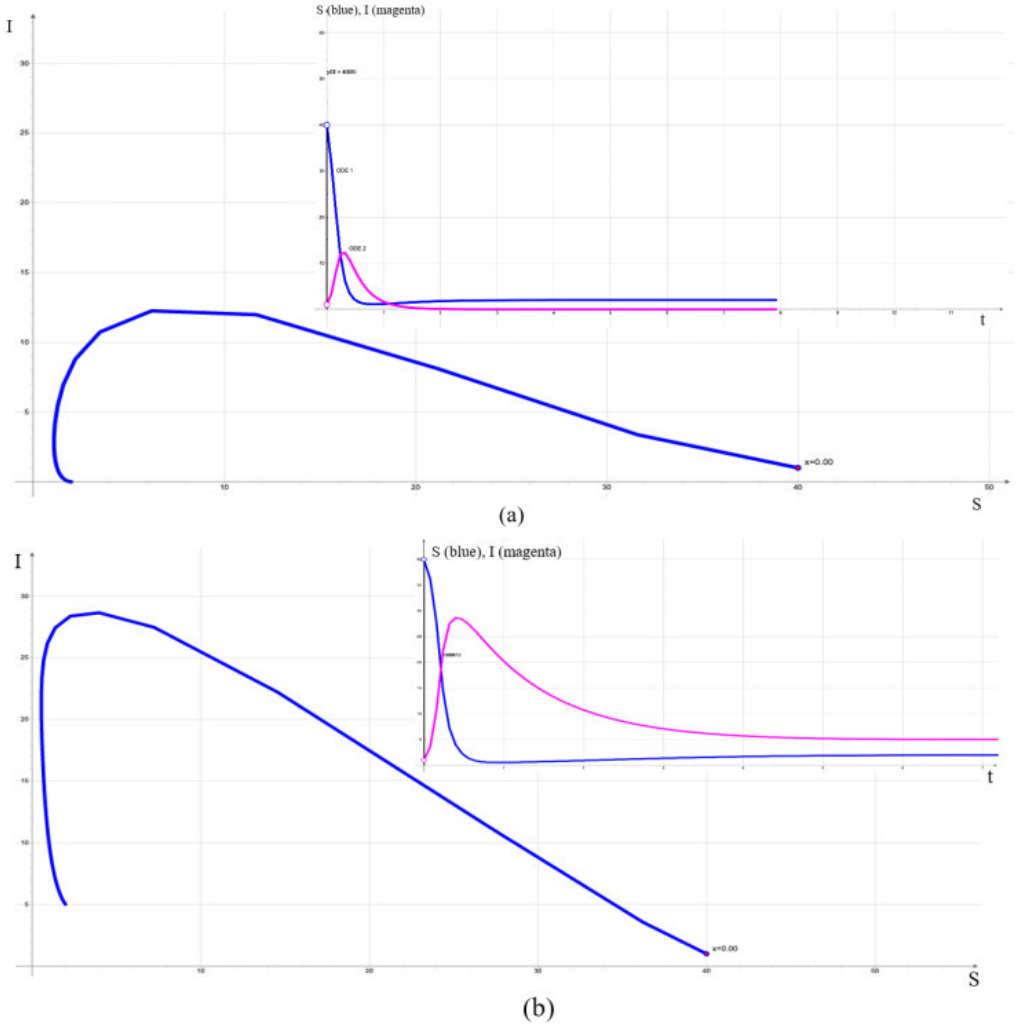


Figure 3.21: Dynamics of the variables S and I in eradication (top) and endemic behavior (bottom) (after https://elsenaju.eu/ Calculator/ODE-System-2-2.htmRWD-1). (a) Eradication $\mathfrak{B} = 0.6, f = 4, \mu = 2, \rho = 0, \sigma = 0, \nu = 4, S(0) = 40, I(0) = 1$; (b) Endemic behavior $\mathfrak{B} = 0.5, f = 6, \mu = 0.5, \rho = 0, \sigma = 0, \nu = 4, S(0) = 40, I(0) = 1$.

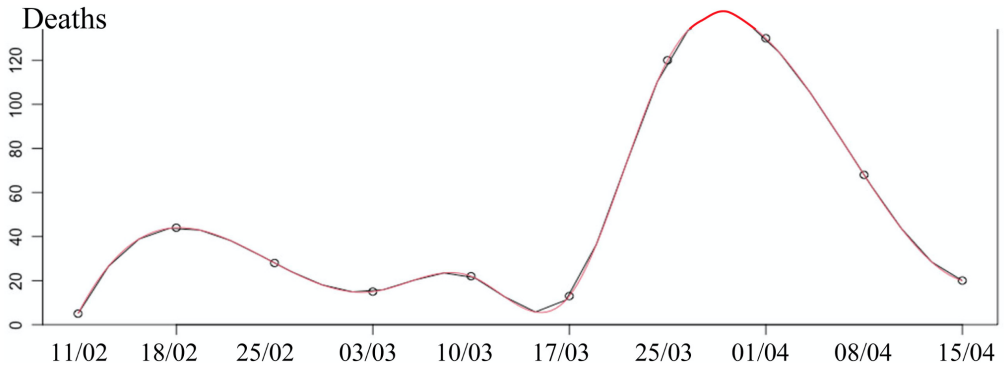


Figure 3.22: Spline graph for COVID-19 deaths in Cameroon during the second wave between week 5 of 2021 and week 14 of 2021.

3.4.1 Mathematical formulation of the identification model

The problem can be formulated in a way that all parameters can be identified from the set $\Theta = \{(S_0, I_0, a, \beta)\} \subset (0, \infty)^4$ if the reported daily infectious cases $I^{'}(t)$ is known for all times $t \geq 0$. It should be noted that it is not sufficient alone to know the value of $I^{'}(t)$ at any time $t \geq 0$, because the knowledge about the value of S_0, the number of susceptible people in the population before the epidemic outbreak is also necessary to allow to obtain a precise information about the values of I_0, a, β (see [137, 153]).
If the solution of the Bernoulli Equation (3.32) with $(d = \nu = 0)$(fertility and mortality rates are negligible):

$$I(t) = \frac{S_0 e^{\beta(t-a)}}{1 + e^{\beta(t-a)}} \text{ and } S(t) = \frac{S_0}{1 + e^{\beta(t-a)}},$$

where S_0, β and a represent respectively the susceptible size at endemic start, transmission rate and endemic start time.
For any time t, we have:

$$I(t) + S(t) = S_0 \left[\frac{e^{\beta(t-a)}}{1 + e^{\beta(t-a)}} + \frac{1}{1 + e^{\beta(t-a)}} \right] = S_0 e^{\beta(t-a)} = \frac{I(t)}{S_0 - I(t)}$$

If the epidemic wave starts at time 0, where $I(0) = 1$, then:

$$1 = \frac{S_0 e^{-\beta a}}{1 + e^{-\beta a}}$$

Then, $S(0) = S_0 - 1$ and a is given by the following equation:

$$a = \frac{\log(S_0 - 1)}{\beta} \tag{3.37}$$

3.4.2 Derivation of the equation of the inflection point

From the Bernoulli's equation, where $\nu = 0$, we have:

$$I^{'}(t) = \frac{dI}{dt} = \frac{\beta}{S_0} SI \times \frac{S_0}{1 + e^{\beta(t-a)}} \times \frac{S_0 e^{\beta(t-a)}}{1 + e^{\beta(t-a)}} = \frac{\beta S_0 e^{\beta(t-a)}}{(1 + e^{\beta(t-a)})^2}$$

The sufficient existence condition for a point of inflection of order 2 for $I^{'}(t)$ in the case that $I(t)$ is 3-times continuously differentiable in a certain neighbourhood of a point t_i, with $I^{'''}(t_i) = 0$, and $I^{''}(t_i) \neq 0$. Then, $I^{'}(t)$ has an inflection point (P_I) of order 2 at t_i, and by differentiating $I^{'}(t)$ twice, then:

$$\begin{aligned} I^{''}(t) &= \frac{d^2 I}{dt^2} = \beta^2 S_0 \frac{e^{\beta(t-a)} - 2e^{2\beta(t-a)}}{(1 + e^{\beta(t-a)})^3} \\ I^{'''}(t) &= \frac{d^3 I}{dt^3} = \beta^3 S_0 \frac{e^{\beta(t-a)}(1 - 6e^{\beta(t-a)} + 2e^{2\beta(t-a)})}{(1 + e^{\beta(t-a)})^4} \end{aligned}$$

Then, the equation given by β and a from its lowest root $x_i = 3 - \sqrt{7} = e^{\beta(t_i - a)} = \frac{I(t_i)}{S_0 - I(t_i)}$ (depending on $I(t_i)$ is:

$$1 - 6x_i + 2x_i^2 = 1 - 6e^{\beta(t-a)} + 2e^{2\beta(t-a)} = 0 \tag{3.38}$$

If $I(0) = 100$, then from the same calculations, the values of β and a can be identified from Equation (3.37) and Equation (3.38).

3.4.3 Identification method and its application

If the value of the time t_i is calculated from the first P_I of an observed wave and if the value of S_0 is known, then it is easy to estimate the maximal reproduction number as $I''(t)$ and identify β and a from Equation (3.37) and Equation (3.38). The steps of identification method are as follows:

- From an interpolation of the curve of the daily reported new cases (see for example red curve in Figure 3.19) of a particular country epidemic wave, calculate the lowest t_i of the wave at which $I'''(t) = 0$,
- From the value of t_i, estimate the maximal reproduction number as $I'''(t_i)$,
- From the knowledge of $S(0)$, the number of susceptible people in the population at the start of a particular country epidemic wave, one can get S_0,
- From Equation (3.37) and Equation (3.38), β and a can be identified.

The identification method is first applied to Cameroon COVID-19 data and then applied to the world COVID-19 outbreak.

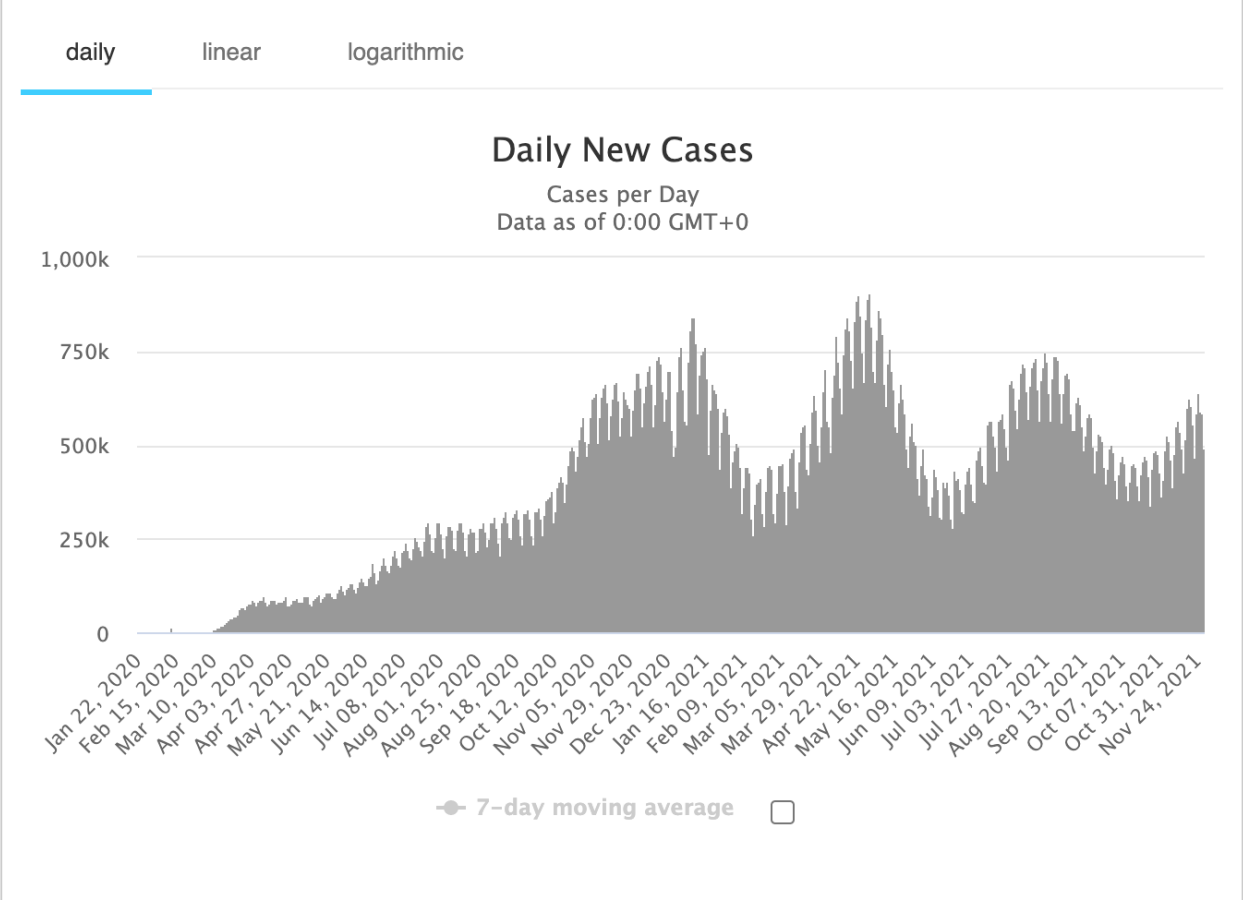


Figure 3.23: Spline graph for COVID-19 deaths in Cameroon during the second wave between week 5 of 2021 and week 14 of 2021.

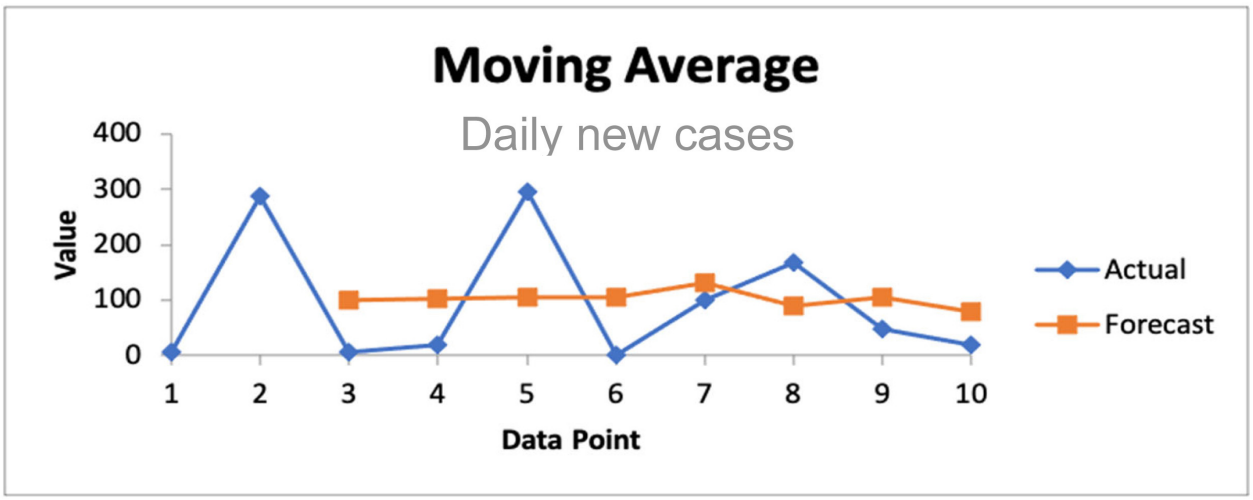


Figure 3.24: Evolution of the daily new cases numbers in Cameroon along the second wave between weeks 1 and 10 of 2021.

For Cameroon, to calculate the time ti of the first inflection point P_I of the second wave of the new infectious cases $X(t)$, for that, the smoothed raw data of the weekly reported new infectious cases in Figure 3.19 by using a cubic spline and the maximal reproduction number is estimated by the maximal value of its slope divided by $X(t)$ at $t = t_i = 4$ (weeks), equal to $R_{0max} = 2.04$, close to the maximal value of the effective reproduction number given for this period by the public database Renkulab [126]. The evolution of the daily new cases and cured cases in Cameroon can be seen in Figure 3.24 and Figure 3.25.

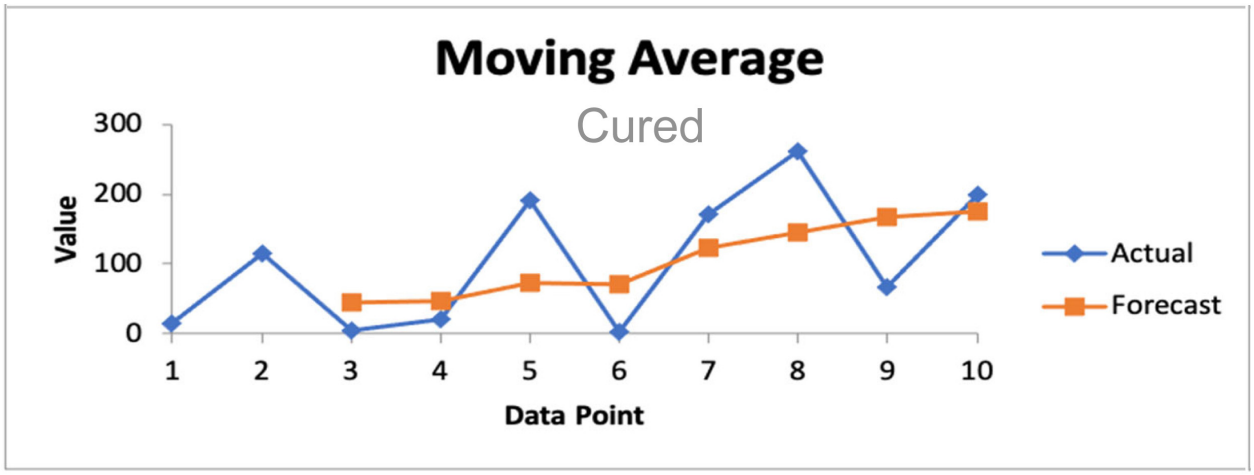


Figure 3.25: Evolution of cured cases numbers in Cameroon along second wave between weeks 1 and 10 of 2021.

From Equation (3.37) and Equation (3.38):

$$a = \log(S_0 - 1)/\beta \text{ and } 1 - 6e^{\beta(t-a)} + 2e^{2\beta(t-a)} = 0,$$

with $S_0 = 27,000,000$ and $t_i = 4$. Then $a = 7/\beta$ and $1 - 6e^{4\beta-7} + 2e^{8\beta-14} = 0$, from which $\beta = 2.0$ and $a = 3.5$, which is consistent for a with the previous estimation of t_i, but overvalued for β, perhaps due to the overestimation of S_0. Indeed, if S_0 corresponds

to the population of the 3 regions (Littoral, Center and Northwest) giving the essential of the infectious cases (for example, $325/371 = 88\%$ of all the new cases during the week of 23/06/2021). More, considering the cumulative number of infectious during the whole week 8 of 2021, there was 2000 cases and if the susceptible population in contact with these infectious is $2000 \times 50 = 100,000$, then from Equation (3.37) $a = 11.5/\beta$, and from Equation (3.38), $\beta = 2.4$ and $a = 4.7$, which shows the sensitivity of calculation of β to initial conditions concerning the susceptible population.

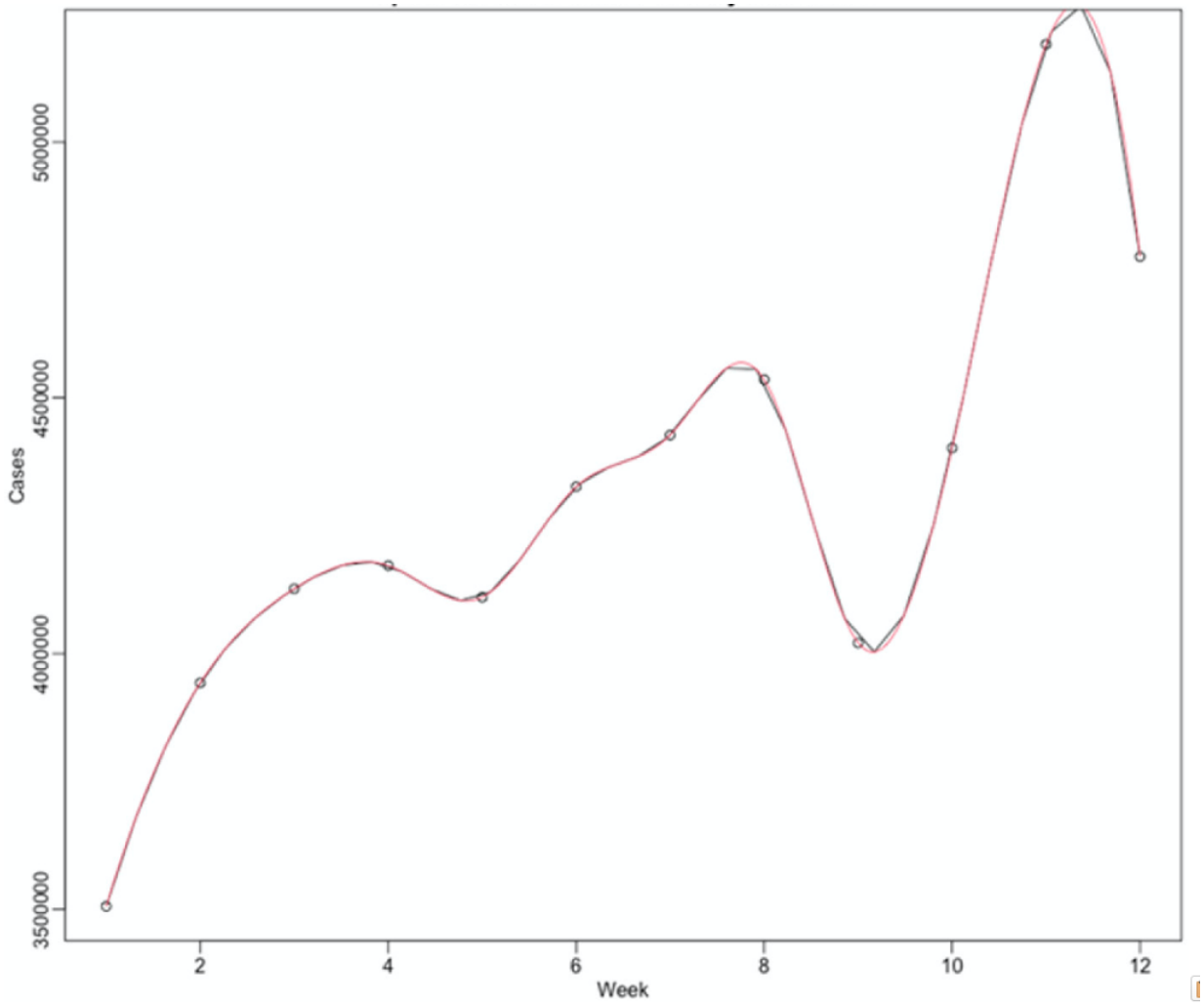


Figure 3.26: Spline graph for COVID-19 deaths in Cameroon during the second wave between week 5 of 2021 and week 14 of 2021.

The same approach is used for COVID-19 outbreak in the world by presenting in Figure 3.23 the global daily new cases of COVID-19 in the world and then smooth the weekly data from October 26 2020 to January 17 2021. The red curve in Figure 3.26 represents the cubic spline approximation of the raw data with the following statistics for the second wave: p-value = 0.05592, Multiple R-squared MR = 0.9949, Relative Mean Square Error RMSE = 41426.05. To calculate the time t_i, of the first inflection point P_I of the second wave (which presents a shoulder and no maximum) of the new infectious cases $X(t)$ for the entire world: this time corresponds to a second wave shoulder, in which the maximum of the new cases is not existing for this wave, hence cannot be used for identifying the epidemiologic parameters as

in [152, 153]. For that, the smoothing of the raw data is done for the weekly reported new infectious cases in Figure 3.23 by using a cubic spline and found $t_i = 4$ (weeks).
Then, from Equation (3.37) and Equation (3.38):

$$a = \log(S_0 - 1)/\beta \text{ and } 1 - 6e^{\beta(t_i - a)} + 2e^{2\beta(t_i - a)} = 0,$$

with $S_0 = 7,000,000,000$ and $t_i = 4$. Then $a = 10/\beta$ and $1-6e^{4\beta-10}+2e^{8\beta-20} = 0$, from which $\beta = 2.575$ and $a = 3.883$, which is consistent with the values of the effective reproduction number calculated in [154].
To conclude this section, methods developed in Section 3.2, Section 3.3 and Section 3.3.2 is applied to Cameroon regions data from [150] in order to obtain the value of the main epidemiologic parameters, using new cases during the beginning of the second wave in 2021 to calculate the discrete R_0 and also the total number of new cases as S_0 in order to calculate the transmission rate β when $t_i = 4$ and finally $R_{\max}$ with $\sigma = 1$ and $\sigma = 2$, corresponding to a duration of the contagiousness period respectively 4 and 8 days. The summary of the result is presented in Table 3.4. The evolution of the daily new cases and cured cases for the regions in Cameroon can be seen in Figure 3.27 and Figure 3.28.
It was observed that there is a big difference between the urban regions, essentially the regions of Yaoundé and Douala (i.e., respectively the Center and Littoral regions) whose urban density is high and the other mainly rural regions.

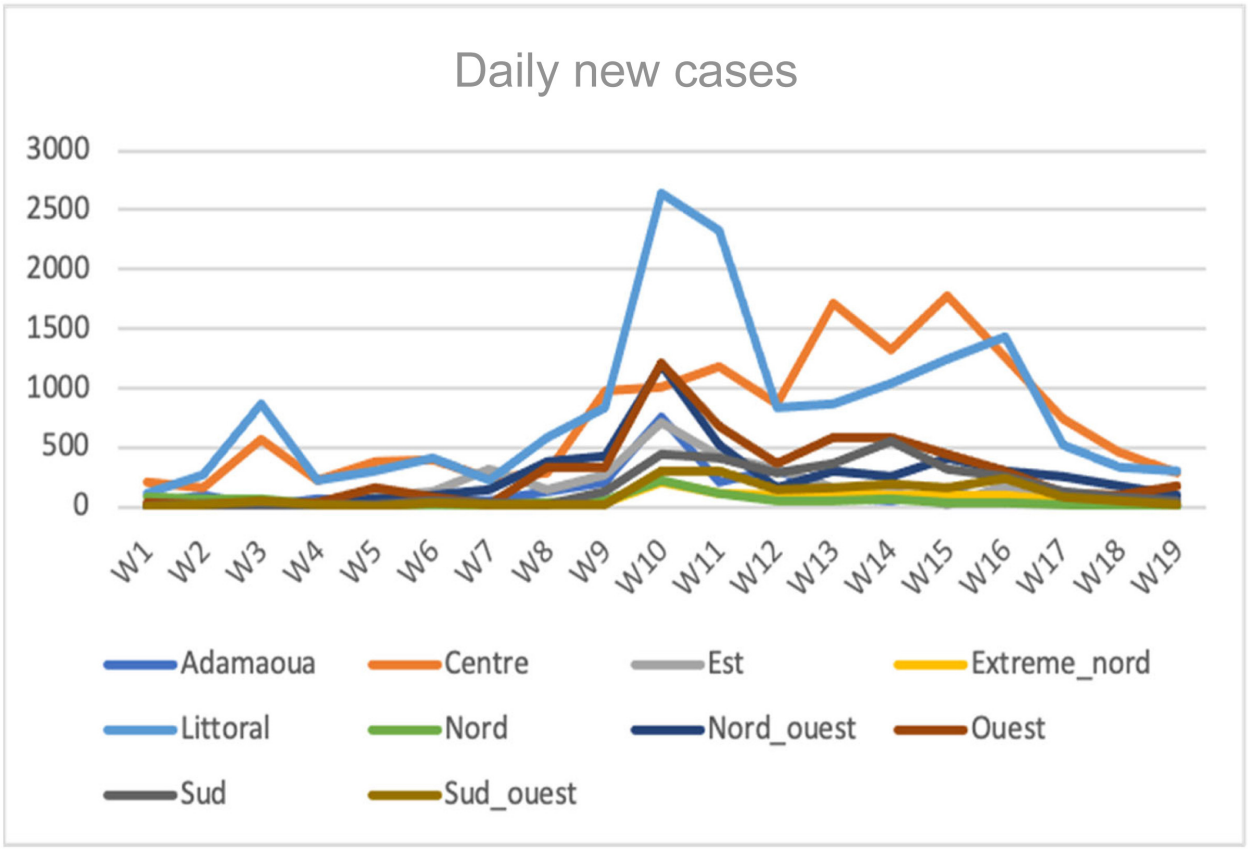


Figure 3.27: Evolution of the daily new cases numbers in the 10 regions of Cameroon along the second wave between weeks 1 and 10 of 2021.

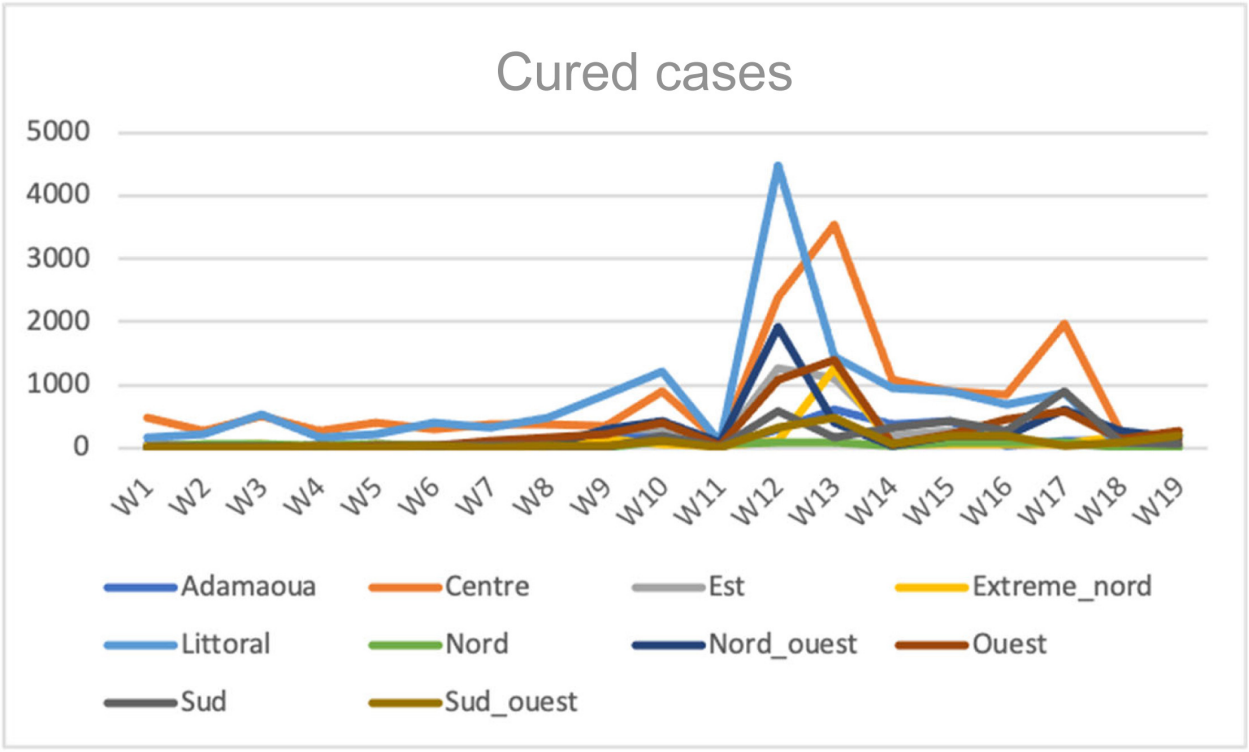


Figure 3.28: Evolution of the cured cases numbers in the 10 regions of Cameroon along the second wave between weeks 1 and 10 of 2021.

Table 3.4: Values of epidemiologic parameters for different regions of Cameroon.

Cameroon regions	**Discrete** R_0	**Transmission rate** β	$R_{\max}(\sigma = 1)$	$R_{\max}(\sigma = 2)$
Adamaoua	1.30175	1.50	2.60143	0.89396
Center	1.04135	1.75	2.85711	0.47851
Est	2.18015	1.50	2.60143	0.89396
Extreme-Nord	1.74880	1.50	2.60143	0.89396
Littoral	0.91404	1.75	2.85711	0.47851
Nord	1.31392	1.50	2.60143	0.89396
Nord-Ouest	3.64468	1.50	2.60143	0.89396
Ouest	3.38611	1.50	2.60143	0.89396
Sud	1.19674	1.50	2.60143	0.89396
Sud-Ouest	0.88279	1.50	2.60143	0.89396

3.5 A Model for the Lifespan Loss Due to COVID-19 Outbreak

3.5.1 Formulation of epidemic ageing model in human populations

Let us consider a population with four age classes ($i = 1, \cdots, 4$) that is $(0, 19)$, $(20, 39)$, $(40, 59)$, and (≥ 60) where f_1, f_2, f_3, f_4 are the fertility rates in the age classes, v_1, v_2, v_3, v_4 are the probability to remain in state i, $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ are the probability to go from state i to state $i+1$, $v_i + \alpha_i = 1 - \mu_i \leq 1, \forall\ i = 1, \cdots, 4$, where μ_i is the death rate at age i. and v_j. The probabilities P_i is on the interval $0 \leq P_i \leq 1 \forall\ i = 1, \cdots, 4$. We can also say that $(v_i + \alpha_i)$ is the general probability of survival. If the two first only being fertile with non-zero fertility rates f_1 and f_2, and if a disease like an epidemic outbreak concerning all the age classes occur adding to the natural mortality rate of each of the three first classes a fatality rate, the sum of the natural and disease dependent mortality, denoted $1 - \alpha_i = \mu_i$. We present in Figure 3.29 the flow chart illustrating the epidemiology ageing model.
The Usher matrix can be written as the following epidemic matrix E_*:

$$E_* = (e_{ij}) = \begin{bmatrix} f_1 + v_1 & f_2 & 0 & 0 \\ \alpha_1 & v_2 & 0 & 0 \\ 0 & \alpha_2 & v_3 & 0 \\ 0 & 0 & \alpha_3 & v_4 \end{bmatrix}$$

The characteristic polynomial of the matrix E_* is given by:

$$\begin{aligned} P(\lambda) &= (f_1 + v_1 - \lambda)(v_2 - \lambda)(v_3 - \lambda)(v_4 - \lambda) - f_2\alpha_1(v_3 - \lambda)(v_4 - \lambda) \\ &= [\lambda^2 - \lambda(f_1 + v_1 + v_2) + v_2(f_1 + v_1) - f_2\alpha_1](v_3 - \lambda)(v_4 - \lambda) \end{aligned}$$

Then, we can calculate explicitly the values of the spectrum of E_*:

$$v_3, v_4, (f_1 + v_1 + v_2) \pm [(f_1 + v_1 + v_2)^2 - 4(v_2(f_1 + v_1) - f_2\alpha_1)]^{1/2}/2$$

The population growth is stable if $\lambda = 1$, increases if $\lambda > 1$ and decreases if $\lambda < 1$.

The proportion of sensitivity s_{ij} of λ (the dominant eigenvalue of E_*) to a variation of the general element e_{ij} of E_*, is given by:

$$s_{ij} = \frac{e_{ij}}{\lambda}\left(\frac{V_i^* V_j^v}{\sum_k V_k^* V_k^v}\right)$$

where V^v is the eigenvector corresponding to the eigenvalue v and V^* is the eigenvector corresponding to λ. Then, the total sensitivity equals 1:

$$\sum_{i,j} s_{ij} = \sum_{i,j} e_{ij} V_i^* \frac{V_j^v}{\lambda V^* . V^v} = \lambda \sum_j V_j^* \frac{V_j^v}{\lambda V^* . V^v} = 1$$

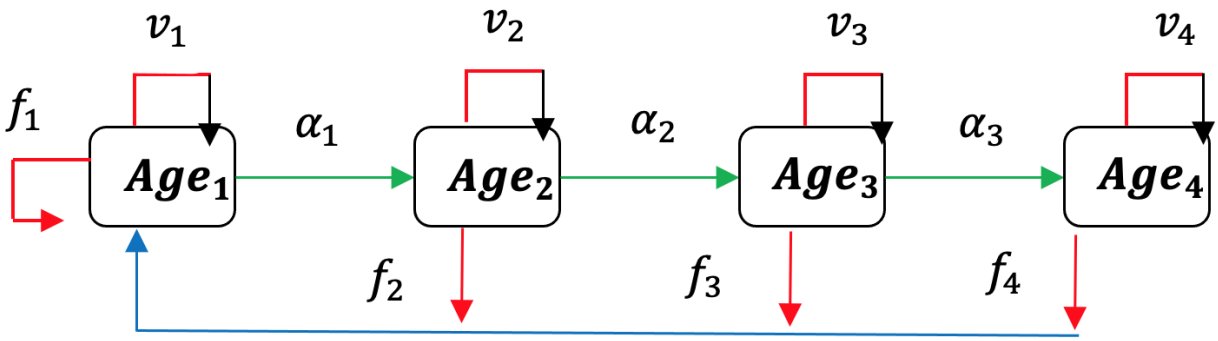


Figure 3.29: Flow chart illustrating the epidemiology ageing model.

3.5.2 Application to COVID-19 outbreak in some countries

We first start by describing the model formulated in terms of normal ageing in Cameroon without considering an outbreak of a viral disease in the population and then apply the model formulated to COVID-19 outbreak in Cameroon, France and Ireland. In the case of Ireland, we extend the age class to five age classes since this country has more data relating to more age groups and most especially the children and teenagers.

Normal ageing in Cameroon

From [155, 156, 157, 158, 159, 160], all the values of parameters like fertility and mortality rates needed to complete the Usher matrix for Cameroon can be found. Table 3.5 gives the list of all the parameters with their signification and Table 3.7 summarizes their values.
Concerning the mortality, the death rate for 1,000 in 2019 is equal to 9.059 [158] and life expectancy at birth is 59.292 years [159]. Values of coefficients f_1, f_2 and f_3 have been calculated in Table 3.7, taking in demographic databases the age-specific fertility rates when available, estimating this rate for the age sub-classes (13,19) and (50,52), considering that the last age class (≥ 60) has no fecundity and weighting all these age-specific fertility rates by the proportions of women in each age class.
Then, considering the values of normal fecundity, ageing and mortality (neglecting infantile mortality), we can calculate the Usher matrix (see Table 3.7):

$$U = \begin{bmatrix} 0.015 + 0.95 & 0.095 & 0.015 & 0 \\ 0.05 & 0.942 & 0 & 0 \\ 0 & 0.049 & 0.93 & 0 \\ 0 & 0 & 0.04 & 0.6 \end{bmatrix}$$

The dominant eigenvalue of U is $\lambda = 1.026$. It is equal to the exponential of the Malthusian parameter of the Cameroon population growth, the real value given in [155] being equal to 1.0258 in 2020.

Table 3.5: List of variables and parameters considered in the paper.

Population dynamics	
$u_i(t)$	Size of the age class i at time t
U, A	Usher matrix, Hahn matrix
f_i	Fertility rate
μ_i	Mortality rate in age i, with $v_i + \alpha_i = 1 - \mu_i$
v_i (resp. α_i)	Probability to remain in age i (resp. go to $(i+1)$) between t and $t+1$
λ (resp. λ, and v)	Dominant eigenvalue (resp. subdominant and current eigenvalue) of U
S, s_{ij}	Sensitivity matrix S and its general coefficient s_{ij}
P_i	Probabilities on the interval $0 \leq P_i \leq 1 \; \forall \; i = 1, \ldots, m$.
u(a,t,s)	Cell concentration at age a, time t and space s
v_i (resp. β_i)	Probability to remain in age i (resp. go to $(i+2)$) between t and $t+1$
Q	Mitotic abortive coefficient $(0 < Q \leq 1)$
Δ (resp. $\square$)	Laplacian (resp. Dalembertian) partial derivative operator
Epidemiologic dynamics	
E_*, e_{ij}	Epidemiologic matrix E_*, its general coefficient e_{ij}
E_M (resp. E_W)	Epidemiologic matrix relative to men (resp. women)
w, H	Invariant measure of E_*, entropy of the invariant measure w of E_*

Application to COVID-19 outbreak in Cameroon

Consider Cameroon, which has a portion of its population affected by the COVID-19 outbreak. The data coming from [119, 155, 156, 157, 158, 159, 160, 161] allow to calculate the epidemic matrix in the cases of normal ageing and supplementary ageing due to the COVID-19 outbreak. We suppose that the fecundity does not change during the epidemy, and we will show the influence of the epidemy during the period of virulence in the host for the infected population. For the sake of simplicity, we suppose that all the infected individuals have the same characteristic of ageing and that the increase of the Malthusian parameter $\log \lambda$, where λ is the dominant eigenvalue of E, concerns a constant percentage of the whole population (equal to the small proportion 2×10^{-4} after [161]).
By neglecting the effect of the viral disease on the fecundity and by taking into account the specific ageing and in worse cases the mortality due to the disease for the fraction of the population affected by COVID-19 [119, 161], with 1418 deaths during the three waves after the first one, between the 1st February 2021 and the 31 January 2022, the matrix E_* becomes with the same reasoning that for the normal population:

$$E_* = \begin{bmatrix} 0.015 + 0.95 & 0.095 & 0.015 & 0 \\ 0.05 & 0.94 & 0 & 0 \\ 0 & 0.044 & 0.8 & 0 \\ 0 & 0 & 0.035 & 0.5 \end{bmatrix}$$

The dominant eigenvalue of E is equal to $\lambda = 1.0236$, slightly less than the observed value 1.0258, and represents the exponential of the Malthusian parameter for the subpopulation affected by the COVID-19, largely less important than the rest of the Cameroon population, which constitutes the large majority of the whole population (99.98 % after [119]).
The data used for Cameroon are all coming from public databases (either demographic or epidemiological) and results from counting published without standard errors. Only median

ages of menarche and menopause have a 95%-Confidence Interval (CI) in the literature (see Table 3.7) and [158, 159]). If we take the left and right limits of these intervals and if we calculate their impact on demographic data, we find for the fertility rate a 95%-CI equal to [0.013, 0.017] at age (0, 19) and [0.014, 0.016] at age (40, 52). Consequently, the dominant eigenvalue of the E_* matrix belongs to the interval [1.02476, 1.02736]. The real value given in [155] is equal to 1.0258 in 2020, and hence, it belongs to this interval.

Application to COVID-19 outbreak in France

We used the same method for calculating and estimating parameters used in Cameroon, which is presented in Table 3.7 for France, while using raw data from France.
From [119, 162], we know that France fertility rate in 2020 is equal to 1.85 children / fertile woman and that women of the second age class (20-39) represents 85% of the whole fertile women during this period of 20 years. Because the second age class size is twice its woman size, the coefficient f_2 of U is equal to ((1.85 / 0.85) / 20) / 2 = 0.108.
The value of the coefficient f_1 (resp. f_3) is obtained in the same way, by considering that only 6% (resp. 10%) of the woman population of the first (resp. third) age class are able to get children. Using data from France [162], population affected by COVID-19 in acute infectious phase is about 4% of the total population at the middle of the fifth wave and mean value of daily cases between 18th and 23rd January 2022 equals 402,984 [119], according to the calculation:

$$\frac{(\text{mean daily cases}) \times (\text{acute phase duration})}{(\text{population size})} = 403,000 \times 7/65,518,000 = 0.043$$

Hence, from Figure 3.30 and from the fecundity rate (1.85) and mortality rate (9.37 / 1000) given in [119, 162], we can calculate with same method as for Cameroon the epidemic matrix for the COVID-19 affected population:

$$E_* = \begin{bmatrix} 0.01 + 0.95 & 0.06 & 0.01 & 0 \\ 0.05 & 0.9 & 0 & 0 \\ 0 & 0.045 & 0.87 & 0 \\ 0 & 0 & 0.035 & 0.5 \end{bmatrix}$$

Then, the dominant eigenvalue of E_* in French COVID-19 population (whose size was 6,100,000 individuals during the month between 19th December 2021 and 18th January 2022) is equal to 1.00107 vs 1.0021 in the general population, and the monthly loss of population due to COVID-19 deaths has been equal to $(1.0021 - 1.00107) \times 6,000,000 = 6,283$, the real observed death number being equal to 127,638 – 121,493 = 6,145.
The coherence between the orders of magnitude of the calculated and observed death numbers in France during the fifth wave constitutes a semi-quantitative validation of the mathematical modelling approach at the population level.

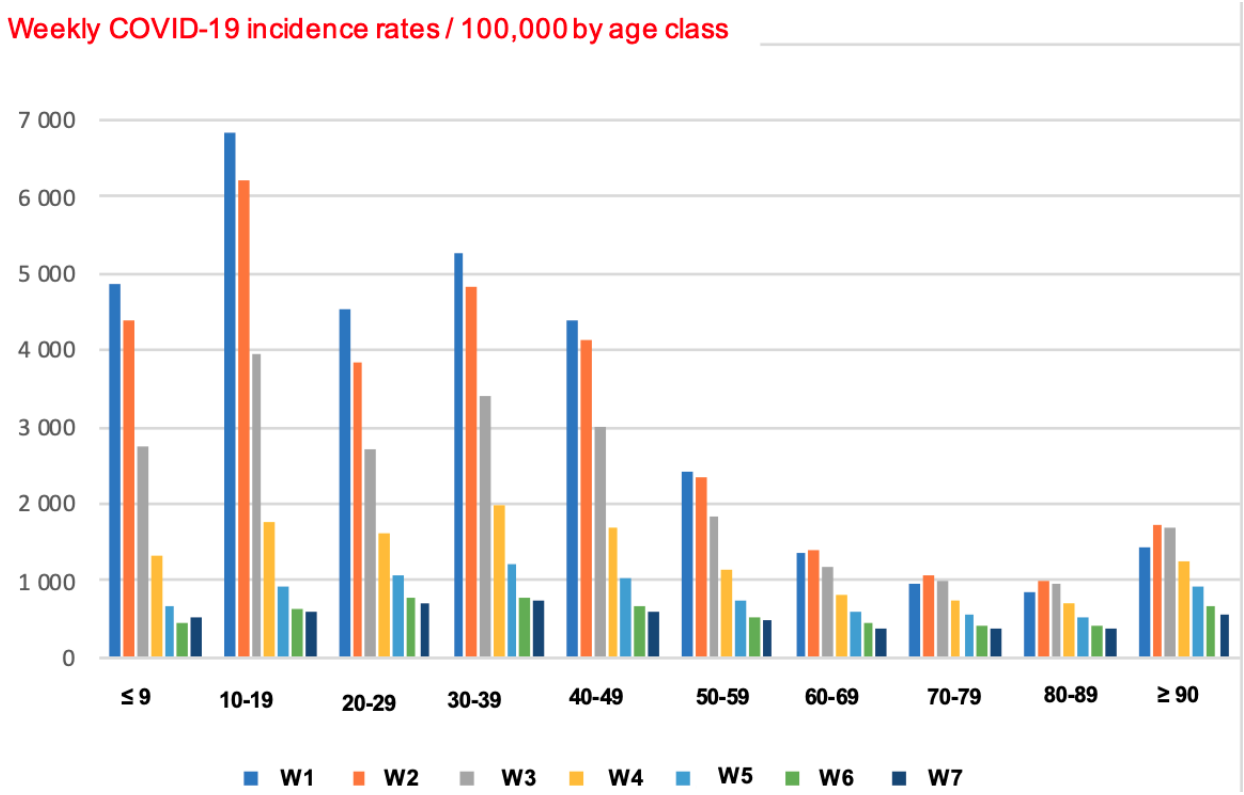


Figure 3.30: Histogram of weekly COVID-19 incidence rates / age class in France during 10 weeks from the 17th January 2022. The global daily incidence rate on the day 17th January 2022 was equal to 3,264 / 100,000 (taking into account the age class sizes) ([162]).

Application to COVID-19 outbreak in Ireland

In Ireland, we will consider five age classes (from estimated data in 2020 [96]) and used the same method for calculating and estimating parameters used in Cameroon, which is presented in Table 3.7, while using raw data from Ireland.

Table 3.6: Distribution of Irish population into 5 age classes in 2020.

Age class	Men	Women
0 - 14	560.338	534.570
15 - 24	316.239	308.872
25 - 54	1098.058	1085.794
55 - 64	278.836	278.498
≥ 65	331.772	383.592

Table 3.6 shows a slight difference due to sex. The fertility rate is equal to 1.808 in 2021 [119]. On Figure 3.32, the epidemiologic data from [163, 164] concerning three young age classes show an increase of notified incidence rate in week 2 of 2022 due to a change in test policy, but a decrease in Epiet data incidence rate at the same time. This can be explained by the fact that the notification requires an administrative validation, which is undoubtedly the cause of the discrepancy observed between the reported incidence curves (on the left in Figure 3.32)

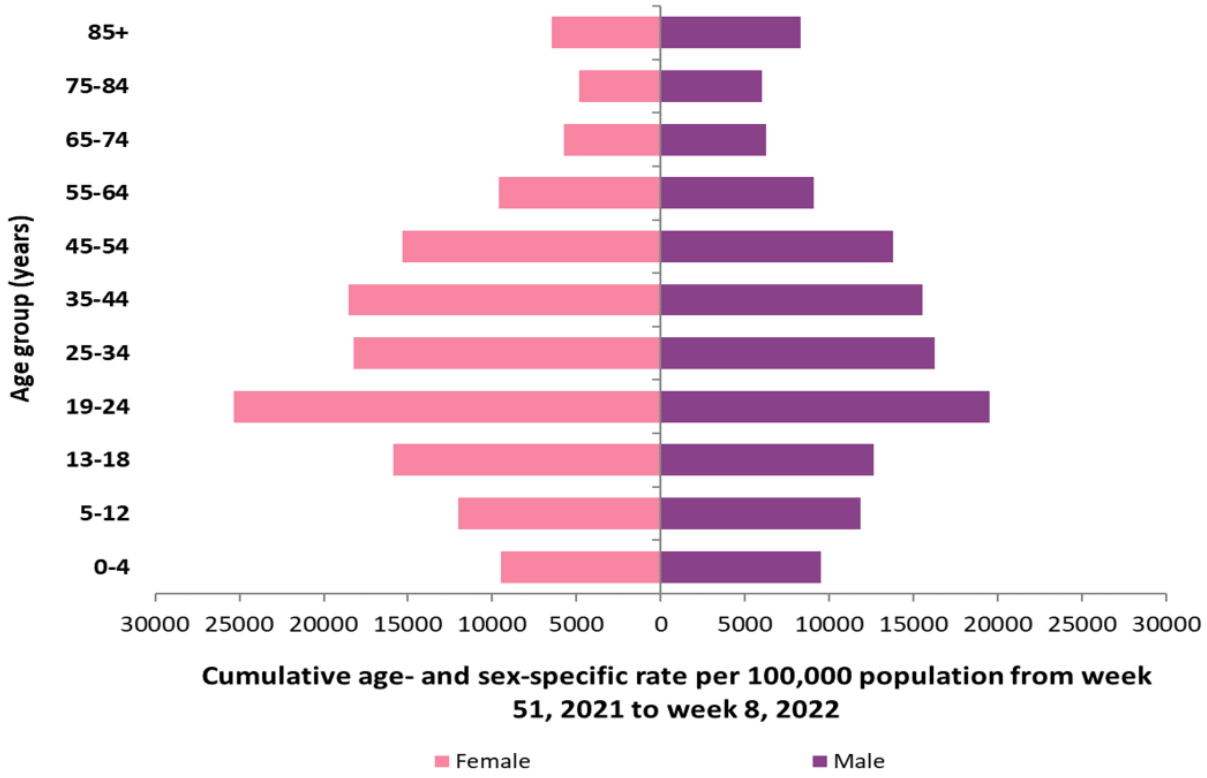


Figure 3.31: Cumulative age and sex-specific incidence rates of confirmed COVID-19 cases per 100,000 population notified in Ireland between week 51 in 2021 and week 8 in 2022 (after [164]).

and the incidence curves early observed by the Irish Health Population Surveillance Center of the European Program for Intervention Epidemiology Training (on the right in Figure 3.32). Figure 3.31 shows a notable difference of incidence rate due to both age and sex.
If we neglect sex influence, the epidemiologic matrix E corresponding to the 5 age classes is:

$$E_* = \begin{bmatrix} 0.01+0.95 & 0.06 & 0.01 & 0 & 0 \\ 0.05 & 0.925 & 0 & 0 & 0 \\ 0.04 & 0.9 & 0 & 0 & 0 \\ 0 & 0.03 & 0.7 & 0 & 0 \\ 0 & 0 & 0 & 0.02 & 0.4 \end{bmatrix}$$

Then, the dominant eigenvalue of E_* is equal to 1.007, which shows that during the month between 19th December 2021 and 18th January 2022, the loss of population due to COVID-19 deaths has been equal to $(1.012 - 1.007) \times 458342/12 = 191$, the real observed death numbers being equal to 6035 – 5835 = 200, which once again confirms the realistic nature of the model.

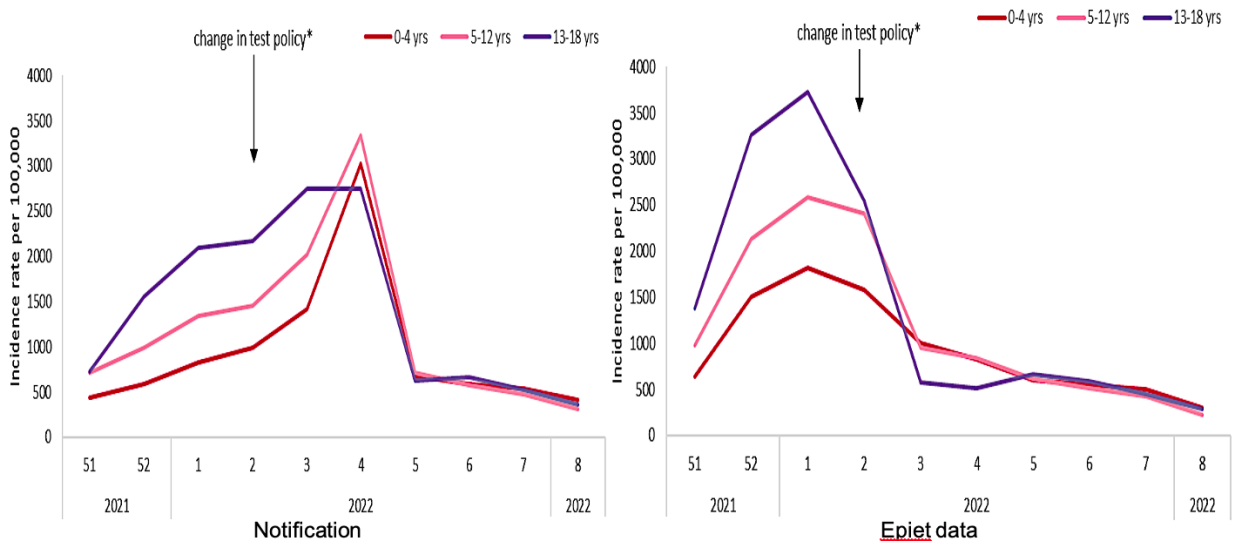


Figure 3.32: Ireland weekly age-specific incidence rates of confirmed COVID-19 cases per 100,000 population among children aged 0 - 18 years by notification (left) and from Epiet data (right) from week 51 in 2021 to week 8 in 2022 (after [163]).

3.5.3 Role of comorbidities

Let us consider now the distribution of age classes of the subpopulation of infected patients presenting the most frequent comorbidity, i.e., cardiovascular pathologies (cf. Figure 3.33) and where males are most numerous than females [119].
For each comorbidity, we can estimate its effect on the Malthusian parameter by taking into account the distribution of the cumulated COVID-19 cases on the age classes. For example, the cardiovascular comorbidity causes the majority of new cases between 40 and 59 years for both sexes, and we have for men (M):

$$E_M = \begin{bmatrix} 0.015 + 0.95 & 0.095 & 0.015 & 0 \\ 0.05 & 0.938 & 0 & 0 \\ 0 & 0.04 & 0.77 & 0 \\ 0 & 0 & 0.03 & 0.4 \end{bmatrix}$$

and for women (W):

$$E_W = \begin{bmatrix} 0.015 + 0.95 & 0.095 & 0.015 & 0 \\ 0.05 & 0.94 & 0 & 0 \\ 0 & 0.043 & 0.79 & 0 \\ 0 & 0 & 0.034 & 0.45 \end{bmatrix}$$

The difference between the values of the exponential growth parameters is equal to that between the dominant eigenvalues: $\lambda_M = 1.0234$ and $\lambda_W = 1.0173$, indicating that among patients suffering from cardiovascular pathologies, men are more affected by the COVID-19 than women, as confirmed by the statistics on 485 cumulated COVID-19 cardiovascular deaths observed among the 22,421 cumulated new cases on 9th September 2020 in Cameroon after 6 months of pandemic [161], from which 278 observed are men and 207 observed are women, that is a sex ratio M / W observed of 1.343 and a differential growth rate ratio calculated equal to (1.026 - 1.0226) / (1.026 - 1.02352) = 1.369.

Table 3.7: List of parameter values used for Cameroon.

Parameter	Nature/Source	Value/Year
Fertility rate of class i	[156, 160]	**Births ‰ women** (2020)
F_1 at age 13 - 14 years	Estimated	[85‰, 95‰]
F_2 at age 15 - 19 years	Raw data	105.8‰
F_3 at age 20 - 24 years	Raw data	211.22‰
F_4 at age 25 - 29 years	Raw data	210.03‰
F_5 at age 30 - 34 years	Raw data	187.84‰
F_6 at age 35 - 39 years	Raw data	138.59‰
F_7 at age 40 - 44 years	Raw data	50.54‰
F_8 at age 45 - 49 years	Raw data	16.57‰
F_9 at age 50 - 52 years	Estimated	[8‰, 12‰]
Age pyramid(P) M,W	[157]	P_{iM}, P_{iW} (2019)
$(0,14) P_{1M}, P_{1W}$	Raw data	21,2%, 20.9%
$(15,19) P_{2M}, P_{2W}$	Raw data	5.4%, 5.3%
$(20,24) P_{3M}, P_{3W}$	Raw data	4.6%, 4.5%
$(25,29) P_{4M}, P_{4W}$	Raw data	4%, 4%
$(30,34) P_{5M}, P_{5W}$	Raw data	3.5% 3.5%
$(35,39) P_{6M}, P_{6W}$	Raw data	2.9%, 2.9%
$(40,44) P_{7M}, P_{7W}$	Raw data	2.3%, 2.3%
$(45,49) P_{8M}, P_{8W}$	Raw data	1.8%, 1.8%
$(50,59) P_{9M}, P_{9W}$	Raw data	2.4%, 2.5%
$(13,14) P'_{1M}, P'_{1W}$	Estimated	3.1% [3.5%, 2.5%]
$(50,52) P'_{9M}, P'_{9W}$	Estimated	0.4% [0.4%, 0.6%]
Median age (menarche)	Raw data (95% -CI) [158]	13.03 [12.47, 13.83] (2016)
Fertility F_1 at age (0,19) = $(F_1 P'_{1W} + F_2 P_{2W})/(P_{1W} + P_{2W})$	Calculated	31.7%
Women % W_1 (0,19)	Calculated	26.2 / 52.8
$f_1 = F_1 W_1$	Calculated	[0.013, 0.017]
Fertility F_2 at age (20,39) = $\sum_{i=3,6} F_i P_{iW} / \sum_{i=3,6} P_{iW}$	Calculated	19.13%
Women % W_2 (20,39)	Calculated	14.9 / 29.8
$f_2 = F_2 W_2$	Calculated	0.095
Median age (menopause)	Raw data (95% -CI) [159]	48 [44, 52] (2005)
Fertility F_3 at age (40,52) = $(F_7 P_{7W} + F_8 P_{8W}) + F_9 P'_{9W}/(P_{7W} + P_{8W} + P_{9W})$	Calculated	29.9%
Women ratio W_3 (40,59)	Calculated	6.6 / 13.1
$f_1 = F_3 W_3$	Calculated	[0.013, 0.017]
Natural death rate / 1000	Raw data [160]	9.059 (2019)
Life expectancy (years)	Raw data [160]	59.292 (2019)
Men / Women ratio	Raw data [119]	59 / 41 (2021)
COVID-19 deaths	Raw data [161]	1418 (2021)
Death rate (20,39) M, W	Raw data	2%,7%
Death rate (40,59) M, W	Raw data	23%,12%
Death rate (>60) M, W	Raw data	34%,22%

Comorbidity	Diabetes		Asthma		Cardiovascular	
Age class	M	W	M	W	M	W
0-4	1	0	0	0	2	0
05-09	0	1	1	1	1	1
10-19	0	3	2	6	1	2
20-29	2	7	5	15	5	10
30-39	15	14	19	44	36	26
40-49	47	28	20	20	42	40
50-59	72	46	15	15	84	56
60-69	70	35	11	9	78	44
70-79	23	12	1	3	23	21
80+	7	3	1	4	6	7
Total	**237**	**149**	**75**	**115**	**278**	**207**

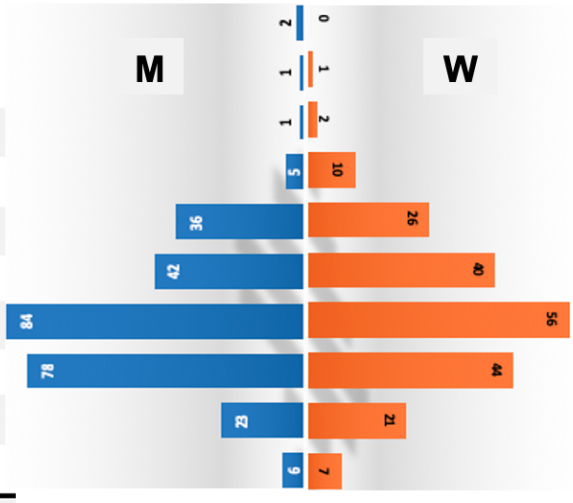


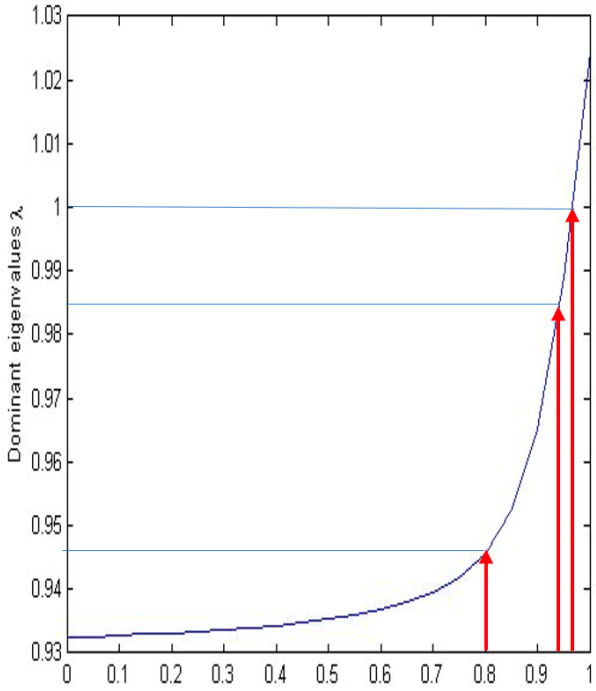
Figure 3.33: On the left, Table of cumulated COVID-19 cases by pathologies and sex on the 9th November 2020. On the right, pyramids of ages by sex for the major comorbidity, the cardiovascular diseases.

3.5.4 Sensitivity analysis and lifespan loss due to a viral disease

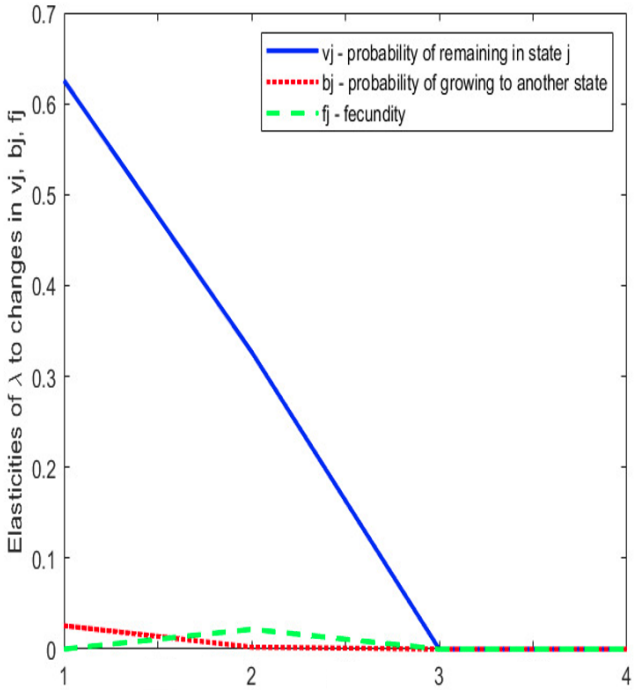
Let consider now as toy example the following epidemiologic matrix of the same form as the previous ones:

$$E_s = \begin{bmatrix} 0.93 & 0.07 & 0 & 0 \\ 0.035 & v_2 & 0 & 0 \\ 0 & 0.03 & 0.7 & 0 \\ 0 & 0 & 0.03 & 0.7 \end{bmatrix}$$

The dominant eigenvalue of E_s is equal to 0.9847, if $v_2 = 0.94$, and 0.9467, if $v_2 = 0.8$. Using the method proposed in [165], Figure 3.34(a) shows dominant eigenvalues while varying value of probability v_2 of remaining in second age class (20-39): when v_2 is greater than 0.96, the dominant eigenvalue is greater than 1 and the population of observed COVID-19 cases exhibits an exponential growth. The second age class is the only class having a large effect on the population growth, because others have no birth rate. Figure 3.34(b) shows the proportion of sensitivity of λ to various changes in ageing, death and fecundity rates.
Figure 3.34(b) confirms that the second age class which contributes the most to population growth is also that which is the most sensitive to its fecundity rate f_2. In Figure 3.34(b), the b_j curve shows also that the probability of remaining in the third and fourth age class does not have effect on λ value while the first age class has the most effect which aligns with our assumption in the equation describing the epidemic ageing model. Biologically, it means that changes of the ageing parameter b in two last age classes do not affect the growth rate of the population. We are also able to deduce from the v_j curve that the largest effect on the sensitivity of λ values is due to the first age class while the second age class has an effect reduced by half and others have no effect, which means that changes in survivability of the first age class is of great importance for the growth of the population.

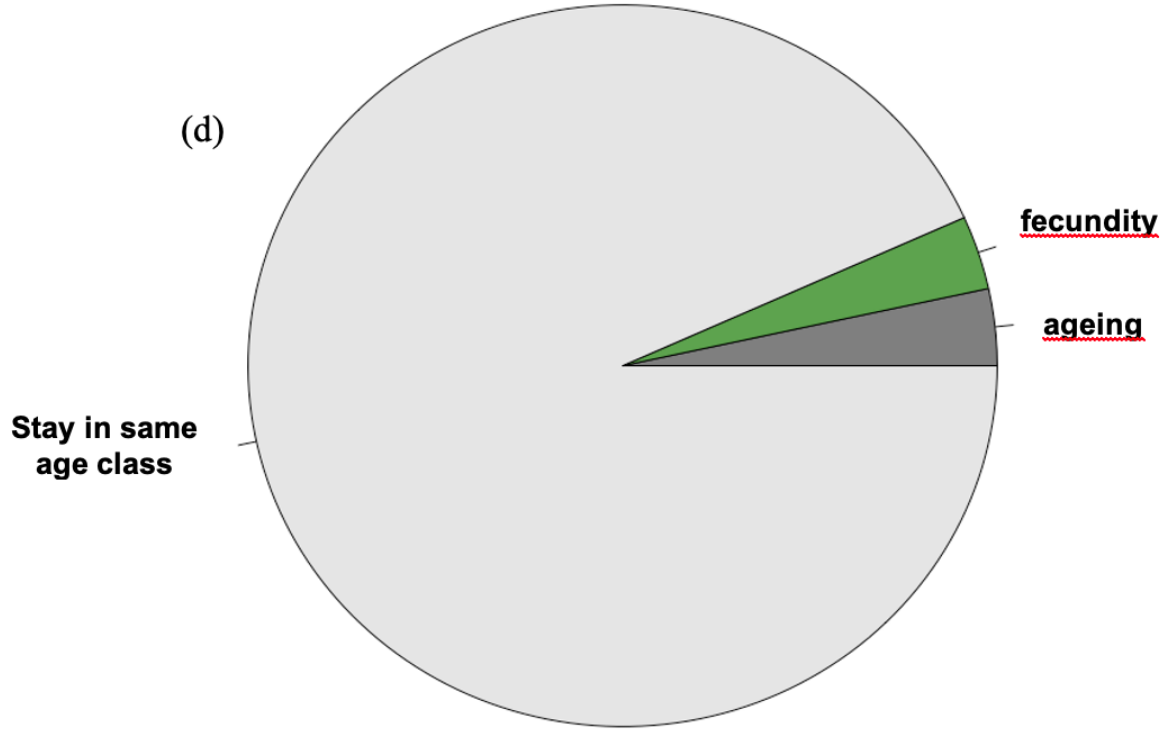


(a) ν_2



(b) Age class j

(d)



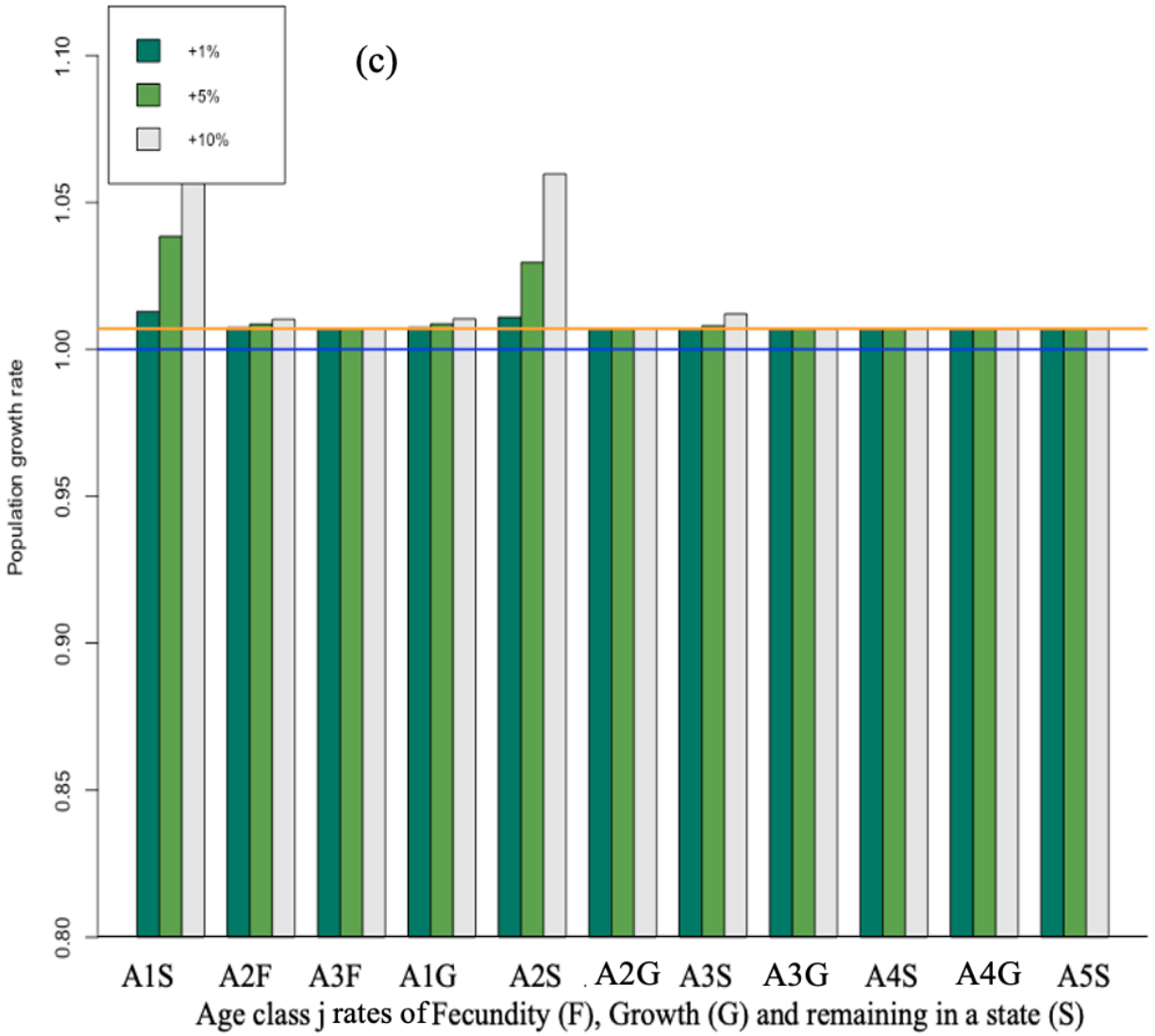


Figure 3.34: (a) Dominant eigenvalue plotted against the varying values of probability v_2 of remaining in the same age class (20-39), during COVID-19 outbreak in France. (b) Sensitivity curves for France. (c) Histogram of changes in dominant eigenvalue λ of E_* for the variations in different percentages of the different E_* coefficients in Ireland. (d) Pie chart to compare the sensitivities for the different categories of E_s coefficients in Ireland.

Figure 3.34(c) affirms that "stay in same state" coefficient of the first two age classes ($v_1 = A1S$ and $v_2 = A2S$) in Ireland is crucial to the population growth which is also the same with the results we observed in France. Figure 3.22d confirms that λ is sensitive to $v_i = AiS(i = 1, 2)$ because corresponding sensitivities are the highest.
The last example of Ireland has shown that it was difficult to overlook the class of young people under 18, who shows an incidence reaching a peak of 3% during the fifth wave (Figure 3.32) and a cumulative rate of approximately 13000 new cases per 100,000 in 10 weeks of this wave (Figure 3.31). Then, we added one more age class in this last example. On the other hand, to refine the coefficients of the epidemiological matrix E_*, it is necessary to better understand the specific ageing processes due to the SARS-CoV-2 virus, in particular that which affects anti-apoptosis proteins such as Gaf1. Indeed, these processes affect the mortality specific to COVID-19, hence the coefficient of the matrix E_* which quantifies the

transition from an age class i to the following age class $i+1$. In what follows, we will seek to lay the foundations for future study of these specific ageing processes.

3.5.5 Cells targeted by SARS-CoV-2 and its implications in cell lifespan loss

In SARS-CoV-2, the main target cells are the alveoli cells of the lungs, the pneumocytes. The turn-over of the 300-700 million of human alveoli [166] is about 4 months [167]. The Type I pneumocytes constitute the major part (95%) of the alveolar surface: they are large (approximately $200\mu m$) and thin (less than $0.2\mu m$) cells, so their barrier to drug transport is at least one order of magnitude lower than typical mucosal or epithelial membranes [168]. Then, each type I pneumocyte covers in mean $5,000\mu m^2$ at the alveolar surface [166]. The diameter of an alveolus is between 200 and $500\mu m$ [169]. If we retain the value of $200\mu m$, their surface is about $280000\mu m^2$, hence we have about 6 pneumocytes/alveolus. For the whole pulmonary tissue, we have then 18×10^8 pneumocytes, from which natural loss each day is about $18 \times 10^8/120 = 1.5 \times 10^6$ cells. Because the SARS CoV-2 is infecting at most 10^7 cells each day [170], the COVID-19 viral disease causes an accelerated ageing in days of the pulmonary tissue, equal to 1.5 times the duration of the acute phase of virulence in the host.

Another origin of specific ageing comes from the hybridization of mRNA of proteins involved in vital metabolisms. In search of hybridization germs, we have inspected viral RNA sequences from different databases [171, 172, 173], using the classic BLAST software. For example, we have already noticed in a previous work [139] that miR 129-5p was a known inhibitor of the biosynthesis of gamma-globin 2, a subunit of human fetal hemoglobin, replaced in adults by beta-globin, also dysregulated in some blood diseases, like the other subunit alpha-globin, by several miRs, including miR 451a [174, 175, 176]. The search for hybridization germs having the same inhibition potential as that of miRs 129-5p and 451a has led to the identification of two subsequences of RNA-dependent RNA polymerase and S genes of SARS-CoV-2. Figure 3.35 shows these subsequences identified as inhibitors of the biosynthesis of human beta-globin. Figure 3.36 shows, for its part, two hybridizations, by subsequences of the gene of the S protein of the Omicron variant of SARS-CoV-2, parts of human interferon mRNA and anti-aging human Gaf1 protein mRNA by subsequences from S protein gene of Omicron variant of SARS-CoV-2 [177].

The primary targets of many viruses are the cells of the most sensitive tissue developing the viral disease, such as in the case of SARS-CoV-2, the cells presenting the ACE2 (for Angiotensin-Converting Enzyme 2) receptor, or in the case of HIV, the cells most infected by the virus, i.e., immune cells CD4+ T cells and macrophages, as well as cells of microglia. In both cases, the viral disease causes a pathologic ageing, even if the patient survives (the death being often due to an opportunistic superinfection). Cytoplasmic nucleases (e.g., RNases) in the cells targeted by the virus are indeed enzymes capable of cleaving the phosphodiester bonds of viral RNA, and the viral genome fragments thus obtained can subsequently form complexes with mRNAs and/or proteins in the host cell, preventing ribosomal translation of proteins, just as miRs do. When targeted proteins are vital, pathogenicity may be greater than that due to viral replication. RNA viruses reproduce their capsid proteins in host cells and duplicate their genome leaving behind RNA fragments, which can behave like miRs in the host genome, if they bind to Argonaut proteins facilitating hybridization to mRNA and then its hydrolysis [178, 179, 180, 181, 182, 183, 184, 185, 186].

In HIV, the main target cells are the T cells of the immune system. HIV can infect up to 2 billion of T cells per day, while no more than 2% (from in mean 500 billion in whole blood of an individual) can be reconstituted per day (by division), that is, about one billion of cells [187]. Then, the HIV virus causes an accelerated aging of the immune system, equal to 2 times the duration of the phase of virulence in the host, which can be chronic in absence of tri-therapy.

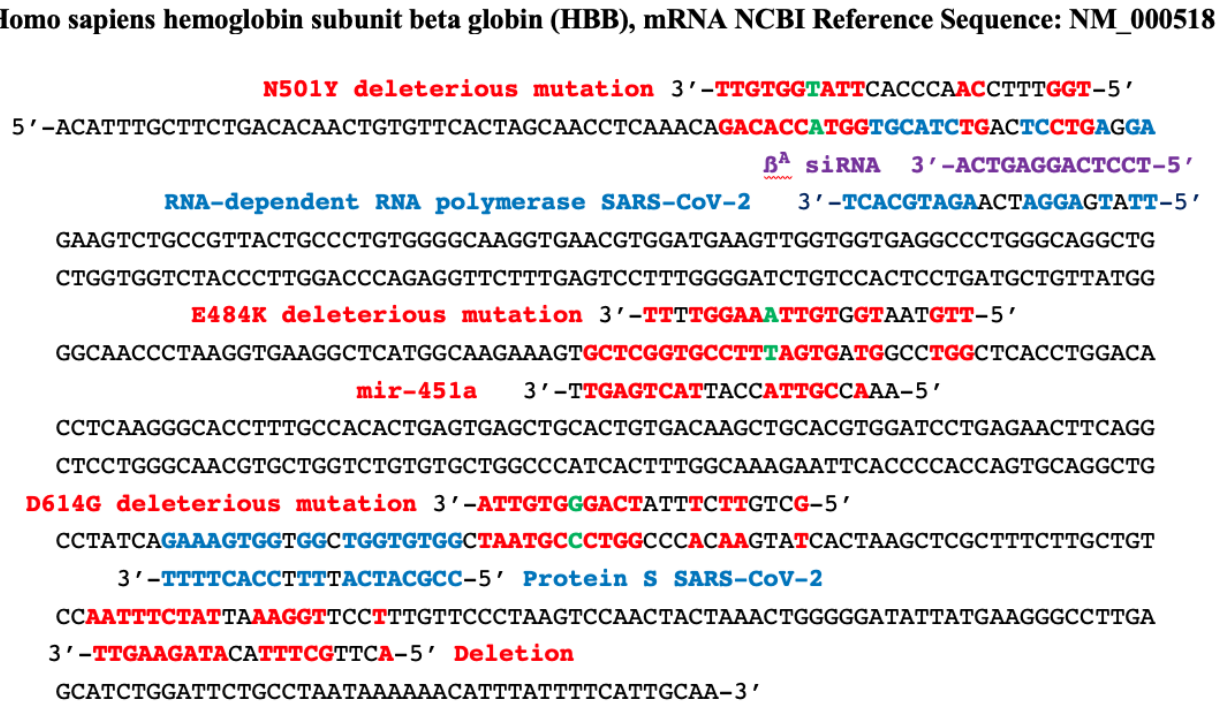


Figure 3.35: Complete fetal human hemoglobin beta (HBB) subunit mRNA sequence potentially targeted by SARS-CoV-2 gene fragments, those from RNA-dependent RNA polymerase (also targeted by the β^A siRNA in violet [187, 188]) and from protein S (in blue), and by the human miR hsa miR 451a (in red). Also shown (in red) are fragments containing deletions of protein S in its N-terminal domain and mutations N501Y, E484K and D614G (base mutated in green). The probability of a hybridization by chance of length 8 in red (resp. 10), for 624 nucleotides, is equal to 0.04 (resp. 0.005), hybridizations TG and GT counting for 1/2.

If we consider the organ level, the high rate of death in COVID-19 patients with cardiac or pulmonary chronic comorbidities (Figure 3.33) indicates that the corresponding organs (heart and lung, respectively) struggle to compensate for the loss of cells destroyed by the SARS-CoV-2 virus, resulting sometimes in a pathologic ageing followed by a failure of these critical organs (critical, because their collapse causes the death of the whole organism of the patient). A source of supplementary pathologic ageing is the inhibition of the biosynthesis of the protein Gaf1, involved in the processes preventing the cell apoptosis, when the viral RNA contains subsequences capable, if it is fragmented by nucleases of the host, to hybridize the mRNA of Gaf1, protein necessary for survive because deeply involved in anti-apoptosis processes [177].
A first example of that is given by fragments of the SARS-CoV-2 virus in Figure 3.36. A second example of the existence of accumulation of small RNA fragments exists in Sclerotinia sclerotiorum infected with the SsHV2-L virus. These virus-derived small RNA fragments

measure about 22 nt, the same length as the miRs, suggesting a cleavage by a Dicer-like protein [189]. Regarding SARS-CoV-2, such an influence on protein translation has already been described [190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202], causing observed effects on the concentration of certain proteins, such as a dramatic decrease in hemoglobin as in other blood diseases [174, 175, 176].

SARS-CoV-2/human/BEL/rega-20174/2021, variant B.1.1.529, GenBank: OL672836.1: 22197-24609 protein S

```
22273 5′-GGTTTCAAACTTTACTTGCTTTACATAGAAGTTATTTGACTCCTGGTGATTCTTCTTCAGGTTGGACAGC
      TGGTGCTGCAGCTTATTATGTGGGTTATCTTCAACCTAGGACTTTTCTATTAAAATATAATGAAAATGGA
      ACCATTACAGATGCTGTAGACTGTGCACTTGACCCTCTCTCAGAAACAAAGTGTACGTTGAAATCCTTCA
      CTGTAGAAAAAGGAATCTATCAAACTTCTAACTTTAGAGTCCAACCAACAGAATCTATTGTTAGATTTCC
      TAATATTACAAACTTGTGCCCTTTTGATGAAGTTTTTAACGCCACCAGATTTGCATCTGTTTATGCTTGG
      AACAGGAAGAGAATCAGCAACTGTGTTGCTGATTATTCTGTCCTATATAATCTCGCACCATTTTTCACTT
      TTAAGTGTTATGGAGTGTCTCCTACTAAATTAAATGATCTCTGCTTTACTAATGTCTATGCAGATTCATT
      TGTAATTAGAGGTGATGAAGTCAGACAAATCGCTCCAGGGCAAACTGGAAATATTGCTGATTATAATTAT
      AAATTACCAGATGATTTTACAGGCTGCGTTATAGCTTGGAATTCTAACAAGCTTGATTCTAAGGTTAGTG
      GTAATTATAATTACCTGTATAGATTGTTTAGGAAGTCTAATCTCAAACCTTTTGAGAGAGATATTTCAAC
      TGAAATCTATCAGGCCGGTAACAAACCTTGTAATGGTGTTGCAGGTTTTAATTGTTACTTTCCTTTACGA
      TCATATAGTTTCCGACCCACTTATGGTGTTGGTCACCAACCATACAGAGTAGTAGTACTTTCTTTTGAAC-3′
```

Homo sapiens Gaf1 family interacting protein 5 (RAB11FIP5), transcript variant 2, NCBI RefSeq: NM_015470.3

```
5′-AGCCTGGGCAAGATGGGCAAAGCCAAAGGCTTCTTCCTCCGCAACAAGCTGCGCAAGTCGTCCCTGACCC
   AGTCCAACACCTCGCTGGGCTCGGACAGCACCCTGTCCTCAGCCAGCGGGAGCTTGGCCTACCAGGGACC
        Variant Omicron Protein S     3′-GTGGTCGACAGGTTGGACTTCTTCTT-5′
   TGGCGCCGAACTCCTCACCCGCTCACCAAGCCGTAGCAGCTGGCTGTCCACTGAAGGGGGCAGGGACTCT
   GCACAGTCCCCCAAGCTGTTCACCCATAAGAGGACCTACAGCGATGAGGCCAACCAGATGCGAGTGGCTC
   CTCCTCGGGCCCTTCTGGACCTTCAGGGCCACCTGGATGCTGCCTCCCGCTCTTCGCTCTGTGTCAATGG-3′
```

Figure 3.36: At the top, partial sequence of the mRNA of the S protein of the Omicron variant of the SARS-CoV-262 virus with indication (in red) of the subsequences of the human interferon gene and the human Gaf1 gene anti- aging that they hybridize. In the bottom, the corresponding hybridization (in red) of part of the anti-aging human Gaf1 gene [177].

If we assume that these short RNA subsequences from the genes of the SARS-CoV-2 virus can bind to Argonaut proteins and hybridize the mRNA of key human proteins involved in important metabolisms such as oxygen metabolism, it follows that mutations and/or deletions observed in the SARS-CoV-2 genome (such as those which appeared in the United Kingdom, South Africa, France or spontaneously in vitro [180, 202]) reinforce the possible existence of these RNA fragments, capable of hybridizing, for example, the mRNA of hemoglobin subunits (Figure 3.35 and Figure 3.36), such as beta-globin [139], impacting oxygen transport in infected patients and of Gaf1 protecting against apoptosis [177]. This mechanism can be marginal but has to be considered in future studies on ageing due to viral infections, which could include a part dedicated to prevention and therapy [203] involving circular RNAs, which serve as “sponges” or “decoys” to small RNA fragments, to prevent them from hybridizing certain proteins vital to the body [204].

3.5.6 Cell ageing due to the virulence: a discrete approach

To model the virus impact on the cell lifetime, causing cell death or disrupting the cell cycle, (when it can be documented), the Hahn model can be used to quantify the loss of function of an organ targeted by the virus. If a certain proportion of cells are destroyed, the virus causes critical organ failure, resulting in the patient's death. Taking the heart as an example, the inhibition of Gaf1 by SARS-CoV-2 (as already proposed in [139]) could activate apoptosis by shortening the cell cycle [177]. Then, using the Hahn model, we could quantify the viral influence on the cell replacement rate in the organ, allowing us to refine the specific mortality rate due to the virus by organ already weakened due to comorbidities. Cell population growth has been already modeled by Hahn [205], using a discrete dynamic ruled by the matrix equation:

$$u(t) = Au(t-1),$$

with Hahn matrix defined by:

$$A = (a_{ij}) = \begin{bmatrix} \alpha_1 & 0 & 0 & \cdots & 2Q\gamma_{n-1} & 2Q\beta_n \\ \beta_1 & \alpha_2 & 0 & \cdots & 0 & 2Q\gamma_n \\ \gamma_1 & \beta_1 & \alpha_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & . & . & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \beta_{n-1} & \alpha_n \end{bmatrix}$$

where $u_i(t)$ represents the size of the cell population in state i of the cell cycle at time t, Q is the mitotic abortive coefficient ($0 < Q \leq 1$), α_i(resp. β_i and γ_i) the probability to remain in state i (respectively to go to state $(i+1)$ and $(i+2)$) between times t and $t+1$, and μ_i mortality rate with:

$$\alpha_i + \beta_i + \gamma_i = 1 - \mu_i \leq 1, \text{ for all } i = 1, \ldots, n.$$

As with the Usher model, the L^2 dynamical stability modulus of the invariant measure w of A is equal to $e^{-|\lambda - \lambda'|}$, where the dominant and sub-dominant eigenvalues of the Hahn matrix if matrix coefficients do not depend on i equal:

$$\lambda = \alpha + (2Q)^{1/n}\beta + (2Q)^{2/n}\gamma \text{ and } \lambda' = \alpha + (2Q\phi)^{1/n}\beta + (2Q\phi)^{2/n}\gamma$$

where ϕ is the second largest absolute value of the n^{th} root of 1. The Kullback-Leibler stability modulus is defined as the cell evolutionary entropy H defined by [142, 206]:

$$H = -\alpha/\lambda \log(\alpha/\lambda) - (2Q)^{1/n}\beta/\lambda \log((2Q)^{1/n}\beta/\lambda) - (2Q)^{2/n}\gamma/\lambda \log((2Q)^{2/n}\gamma/\lambda)$$

In certain cases, cell therapies could cause a rejuvenation of the tissue damaged by the virus, and the value of $\alpha(a)$ could in this case increase and change the value of H.

3.5.7 Cell ageing due to the virulence: a continuous approach

Several works have introduced continuous models with the demographic variable age in order to differentiate the reactions to the virulence of the different age groups of a population, in particular because of an immune response which gradually decreases with age [207, 208, 209].

We consider also as an ultimate perspective the building of a global continuous model integrating both the cell cycle of the organ cells and the age of the patient. For that purpose, we recall that the continuous equivalent of Usher model derives from the classical von Foerster equation [210], where u(a,t,s) is the cell concentration at age a, time t and space s:

$$\frac{\partial u}{\partial a} + \frac{\partial u}{\partial t} = -\mu(a)u$$

By adding a second order term $\frac{\partial^2 u}{\partial a^2}$ taking into account the existence of the parameter γ in the Hahn matrix, the dynamical behavior of this continuous equation is the same as for the discrete Hahn model. The main interest of the continuous formulation is the possibility to add a diffusion term [211], if cells are moving to repair an organ damaged by the virus:

$$\frac{\partial u}{\partial a} + \frac{\partial u}{\partial t} - (a)\frac{\partial^2 u}{\partial a^2} - \sigma(a)\Delta u = -\mu(a)u \Leftrightarrow \frac{\partial u}{\partial a} + \frac{\partial u}{\partial t} - \Box u - = -\mu(a)u,$$

where $\Box u = \frac{\partial^2 u}{\partial a^2} - \sigma(a)\Delta u$ can be considered as a Dalembertian operator taking into account accelerated ageing and diffusion in space. Then, we obtain the most general continuous operator including pathological cell ageing and cell motion represented by what J. Besson and J.P. Caubet [212, 213] called the charge of Sinbad the Porter (the Dalembertian symbol $\Box$) and the sail of Sinbad the Sailor (the Laplacian symbol Δ).

3.6 Age Dependent Epidemic Modeling of COVID-19 Outbreak

The purpose of this section is to improve age dependent modeling which was presented in the previous section by using a deterministic approach with a system of ordinary differential equations and we have been able to also show in the previous section the influence of the median age of a country on the incidence of COVID-19 which has been highlighted, in particular through the appearance of many asymptomatic cases in the age groups below 50.For example, on the site of Johns Hopkins University dedicated to COVID-19 [214], data clearly shows this influence on the case fatality rate corresponding to the cumulative deaths recorded 5 months after the beginning of the outbreak (12 May 2020) vs. the median age of many countries in 2017 (Figure 3.37). This first observation has been confirmed by studies in France (Figure 3.40), and we will confirm in this section that the age of the patients suffering from COVID-19 is a good predictor of severity.
The data used in this section is from public epidemiologic and demographic databases [157, 215] and the reference methods are both from classical demographic (such as Leslie) and epidemiologic (such as Ross and Kermack-McKendrick) models. In Figure 3.38 and Figure 3.39, we present visualizations for pandemic dynamics in different age groups and sexes for Kuwait and Cameroon, respectively, to support the motivation for this section that pandemic evolution and severity are related to age classes, and thus modeling is important as various researchers and health experts are investigating SAR-CoV-2 mutations. The countries under consideration have a higher proportion of young people, while the elderly have a lower proportion. Only 2% of the total population in Kuwait is over 65 and vulnerable to the pandemic, whereas 20.8% of the population in France is over 65 and vulnerable to the pandemic, and 2.7% of the population in Cameroon is over 65 and vulnerable to the pandemic.

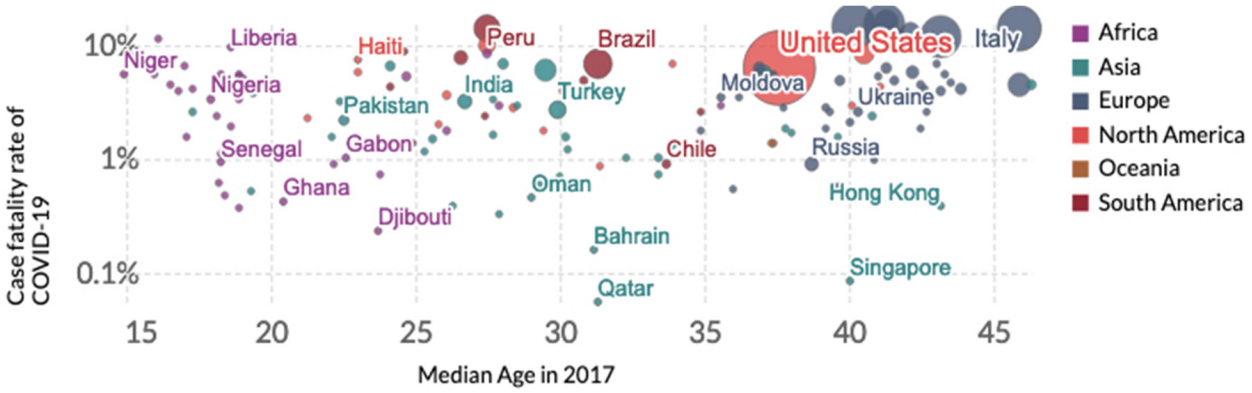


Figure 3.37: Dependence of the case fatality rate (from cumulative deaths on the 20 May 2020) vs. median age of several countries in 2017 (from [214]). The area of a country circle is proportional to the number of cumulated deaths due to COVID-19 on the 20 May 2020, e.g., for the USA: 99,643 (in red).

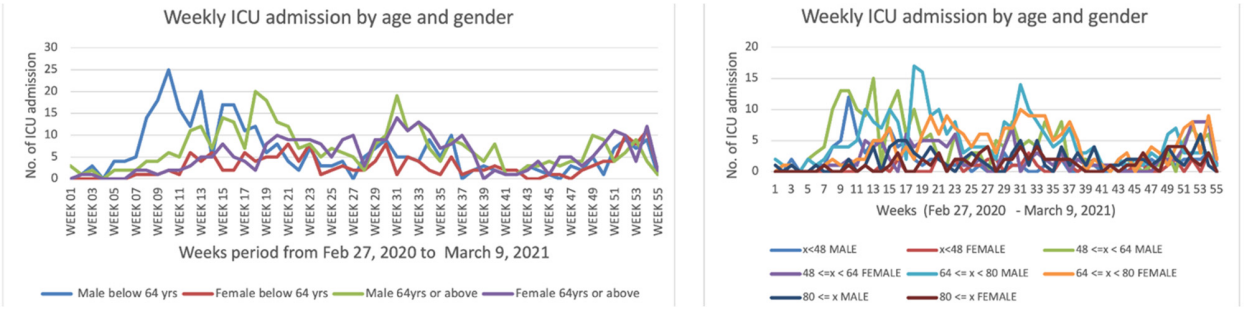


Figure 3.38: Curves of COVID-19 weekly ICU admissions in Kuwait by age classes (two on the **left** and four on the **right**) and gender from 27 February 2020 to 9 March 2021.

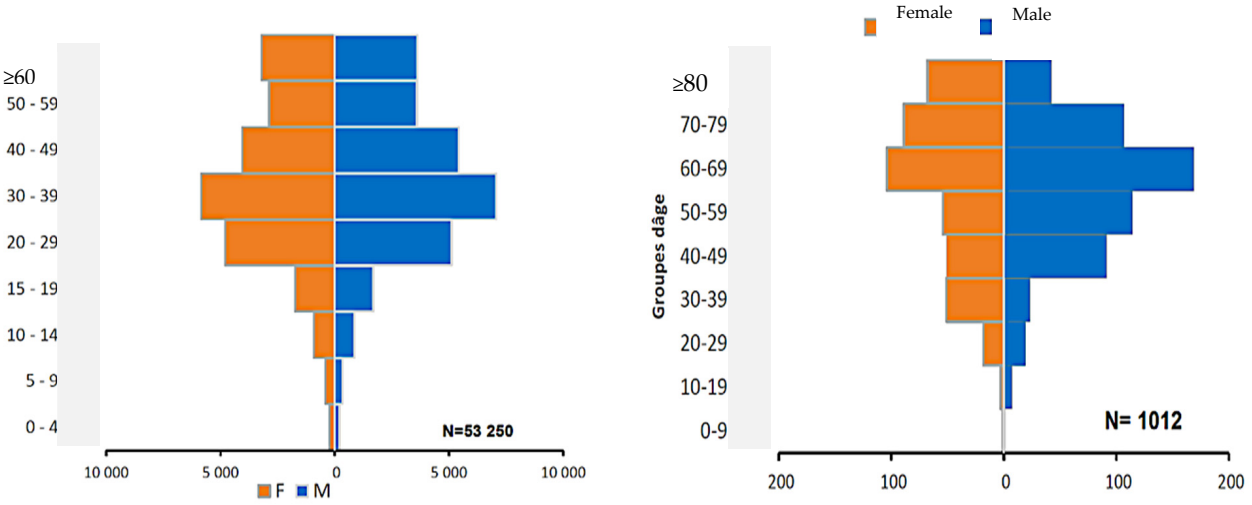


Figure 3.39: **Left**: Distribution of cumulated confirmed cases of COVID-19 by age group and gender in Cameroon as of 23 June 2021 [216]. **Right**: Distribution of deaths due to COVID-19 infection by age group and gender in Cameroon as of 23 June 2021 (after [150]).

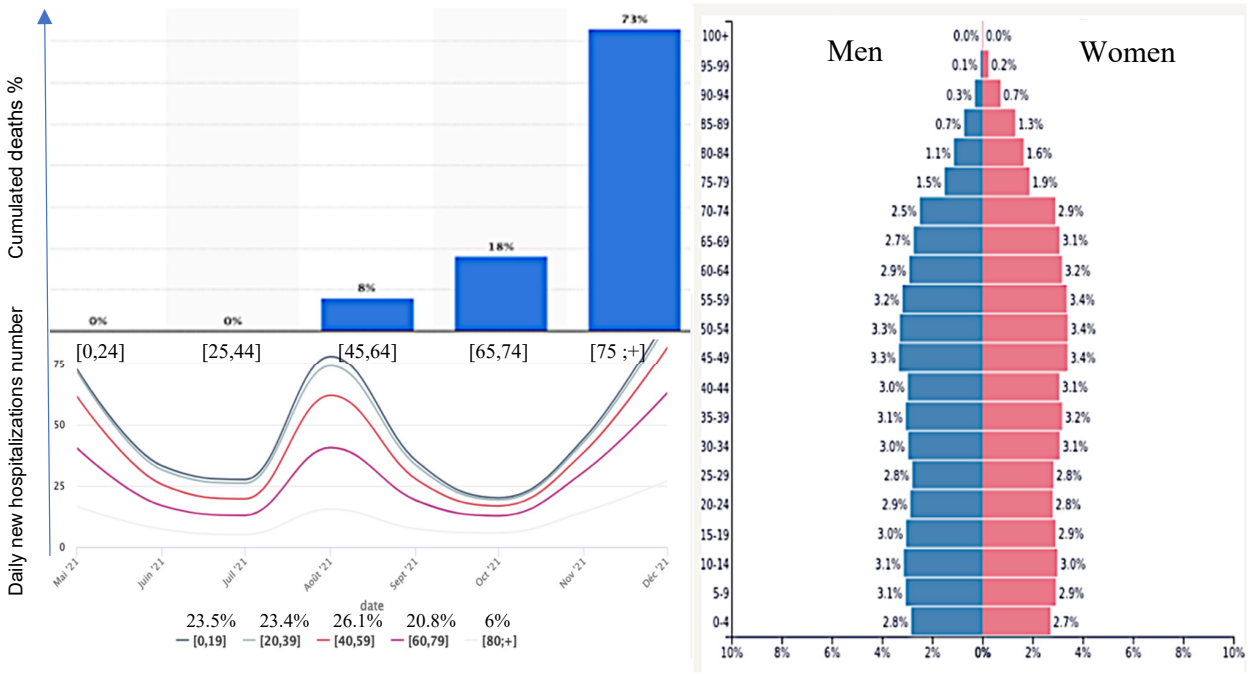


Figure 3.40: **Top left**: COVID-19 percentage of death in France by age class [215]. **Bottom left**: Influence of age (curves with color coding) on COVID-19 hospitalizations in France in the extreme age classes [215]. **Bottom right**: age classes pyramid in 2020 in France (total population size: 65,273,512) [157].

We propose a model which is an improvement of the Ross and Kermack-McKendrick model by trying to compensate some deficiencies in this model. We will first introduce two age classes to account for adults and the elderly in the population and then take account of vaccination before applying our model to some countries chosen as examples.

3.6.1 New SIGR model formulation

The heterogeneity in the age of the populations studied here (Kuwait, France, and Cameroon) leads to considering a model in age groups with specific susceptibilities for each age class, which makes the prediction problem of the new infectious growth more difficult. Basically, there are three age groups of interest in the COVID-19 outbreak, which are, respectively, 0–19 years, 20–64 years, and >64 years, but here we only consider two (20–64 years and >64 years) age groups because the group 0–19 years is widely seen as being less infectious by SARS-CoV-2 since this age group had a low infection rate throughout the period considered in this study, especially the countries data used for simulation.
This heterogeneity in individual age and the reaction to biological and environmental changes that have been observed in COVID-19 dynamics in terms of different reactions to vaccination by age group, severity of infection per age group, hospitalization, and intensive care unit (ICU) records show different patterns, which is why it is important to improve mathematical models for COVID-19 pandemic prediction to account for different proportions of ages in the population, which is a major factor in epidemic history [216]. Here, the originality is to propose an improvement of the classical Ross and Kermack-McKendrick model and use it for giving a theoretical and numerical framework for interpreting the relationship between

demographic parameters such as age pyramid, fertility and mortality, and epidemiological parameters such as the basic reproduction number R_0 and vaccination rate.

Table 3.8: List of the parameters considered in the SIGR (Susceptible–Infectious–Goneanewsusceptible–Recovered) model.

Epidemiologic Parameters	
S_i	Susceptible individuals of age class i
I_i	Infectious individuals of age class i
G_i	Gone anew susceptible individuals of age class i
R_i	Fully recovered and resistant individuals of age class i
R_0	Basic reproduction number
ϕ_i	Transmission rate of age class i
η_i	Relapsed rate of age class i
γ_i	Recovering rate of age class i
a_1^K	Survival rate from age 1 to age 2 for compartment K of age class 1
μ_2^K	Natural death rate of compartment K of age class 2
Demographic Parameters	
β_{i1}	Birth & loss of resistance of recovered from compartments of age class i
β_{22}	Loss of resistance of recovered rate from compartments of age class 2
θ_i^K	Vaccination rate from compartment K of age class i
ϵ_i	Specific fatality rate ϵ due to the disease of age class i

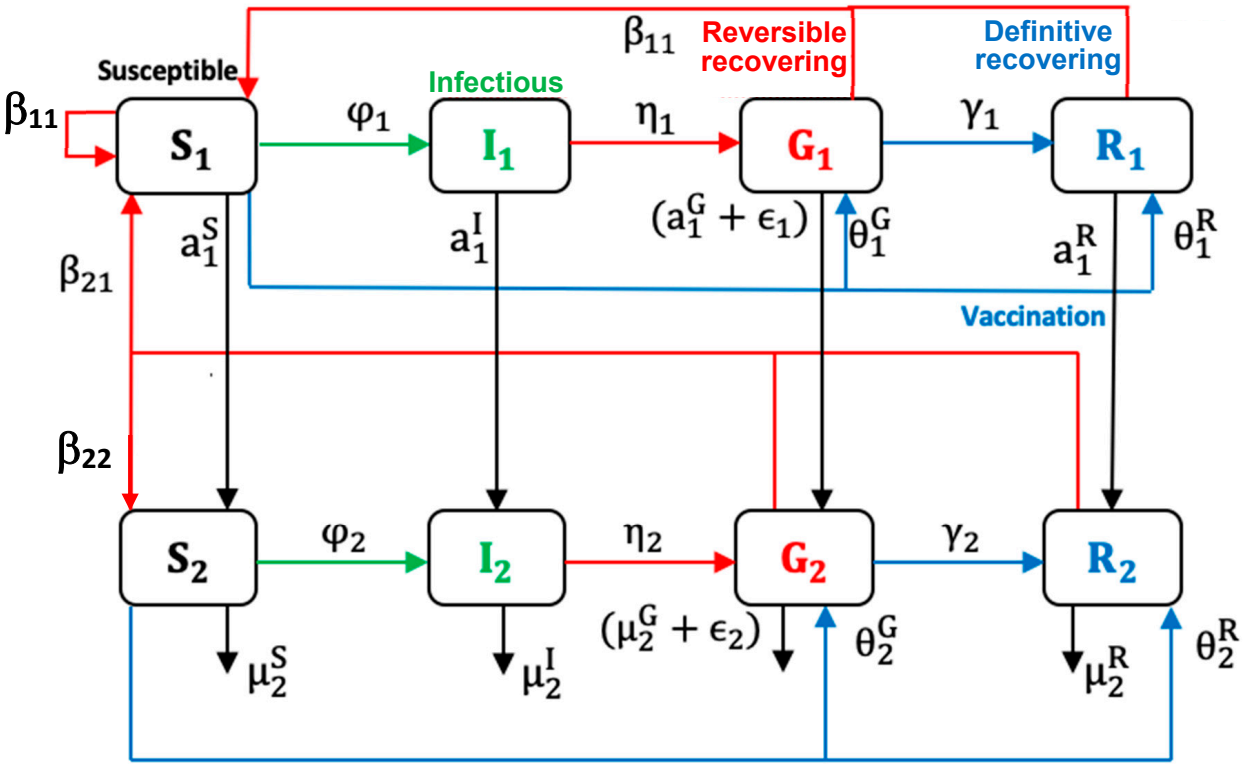


Figure 3.41: Age-dependent scheme for COVID-19 outbreak modeling.

We propose a Susceptible–Infectious–Goneanewsusceptible–Recovered (SIGR) model as an

improvement of the Ross and Kermack-McKendrick models including age class and vaccination state for COVID-19 in a given population. Neglecting differences between kids and young adults, we only retain two age classes: adult and elderly, 1 and 2. We assume that all infectious are symptomatic and we consider birth and natural death rates β and μ, as well as specific fatality rate ϵ due to the disease. Age groups $(i = 1, 2)$ concern individuals susceptible, infectious I_i, gone anew susceptible G_i, and a fully recovered and resistant R_i. We denote for each age group $(i = 1, 2)$ the transmission rates φ_i, fertility, and loss of resistance rates β_{ij} (supposed to be equal inside an age class, for the sake of simplicity), natural death rates $\mu_2^S,\ \mu_2^I,\ \mu_2^G$ and μ_2^R, vaccination rates group θ_i^G and θ_i^R, survival rates from age 1 to age 2 $a_1^S,\ a_1^I,\ a_1^G$ and $a_1^R,$ specific death rate due to the disease ϵ_i, relapsed rate η_i and recovery rate γ_i (cf. Table 3.8).
The description above can be illustrated by the following set of non-linear differential Equation (3.39), while the graphical representation of the model is given in Figure 3.41:

$$\begin{aligned}
\frac{dS_1}{dt} &= \beta_{11}\ (S_1 + G_1 + R_1) + \beta_{21}\,(S_2 + G_2 + R_2)\ -\ \left(a_1^S + \theta_1^G + \theta_1^R + \varphi_1(I_1\ + I_2)\right)S_1 \\
\frac{dS_2}{dt} &= a_1^S S_1 - {\mu_2}^S S_2\ -\ \left(\theta_2^G + \theta_2^R + \varphi_2(I_1 + I_2)\right)S2\ + \beta_{22}\,(G_2 + R_2) \\
\frac{dI_1}{dt} &= \varphi_1\,(I_1 + I_2)\,S_1 - (a_1^I + \eta_1)I_1 \\
\frac{dI_2}{dt} &= \varphi_2\,(I_1 + I_2)\,S_2 + a_1^I I_1 - \eta_2\ I_2 - {\mu_2}^I I_2 \\
\frac{dG_1}{dt} &= \eta_1 I_1 + \theta_1^G S_1 - \left(a_1^G + \gamma_1 + \epsilon_1\right)G_1 \\
\frac{dG_2}{dt} &= \eta_2 I_2 + a_1^G G_1 + \theta_2^G S_2 - \left(\gamma_2 + {\mu_2}^G + \epsilon_2 + \beta_{22}\right)G_2 \\
\frac{dR_1}{dt} &= \gamma_1 G_1 + \theta_1^R S_1 - a_1^R R_1 \\
\frac{dR_2}{dt} &= \gamma_2 G_2 + \theta_2^R S_2 + a_1^R R_1 - ({\mu_2}^R + \beta_{22})R_2
\end{aligned} \tag{3.39}$$

with $S_1(t) \geq 0, S_2(t) \geq 0,\ I_1(t) \geq 0,\ I_2(t) \geq 0,\ G_1(t) \geq 0,\ G_2(t) \geq 0,\ R_1(t) \geq 0,\ R_2(t) \geq 0.$

3.6.2 Positivity and boundedness of the solution

Lemma 1. *Let the initial conditions be given as follows:*
$\{S_1(0), S_2(0),\ I_1(0),\ I_2(0),\ G_1(0),\ G_2(0),\ R_1(0),\ R_2(0) \geq 0\}$, *then solutions of the system of equations (3.39) are positive for all* $t > 0$.

Proof. From the first equation in the model Equation (3.39), we obtain:

$$dS_1/dt = \beta_{11}\ (S_1 + G_1 + R_1) + \beta_{21}\ (S_2 + G_2 + R_2)\ -\ \left(a_1^S + \theta_1^G + \theta_1^R + \varphi_1(I_1 + I_2)\right)S_1,$$

Hence, $dS_1/dt \geq \beta_{11}\ - a_1^S - \theta_1^G - \theta_1^R)S_1$.
By using the separating variable method and then, integrating, we obtain:

$$\int \frac{dS_1}{S_1}\ \geq \int \left(\beta_{11}\ - a_1^S - \theta_1^G - \theta_1^R\right)dt$$

$$\ln S_1 \geq \left(\beta_{11} \ - a_1^S - \theta_1^G - \theta_1^R\right) t + k$$

Finally, by writing $S_1(0) = e^k$, we have:

$$S_1(t) \geq S_1(0)\, e^{(\beta_{11} - a_1^S - \theta_1^G - \theta_1^R)t} \geq 0.$$

By applying the same process to other equations in (1), we have:

$$\begin{aligned}
S_2(t) &\geq S_2(0)\, e^{(a_1^S - \theta_2^G - \theta_2^R - {\mu_2}^S)t} \geq 0 \\
I_1(t) &\geq I_1(0)\, e^{-(a_1^I + \eta_1)t} \geq 0 \\
I_2(t) &\geq I_2(0)\, e^{-(\eta_2 \ I_2 + \ {\mu_2}^I)t} \geq 0 \\
G_1(t) &\geq G_1(0)\, e^{-(a_1^G + \gamma_1 + \epsilon_1)t} \geq 0 \\
G_2(t) &\geq G_2(0)\, e^{-(\gamma_2 + \ {\mu_2}^G + \epsilon_2)t} \geq 0 \\
R_1(t) &\geq R_1(0)\, e^{-a_1^R t} \geq 0 \\
R_2(t) &\geq R_2(0)\, e^{-{\mu_2}^R t} \geq 0
\end{aligned}$$

Then, the solutions of the system of equations (3.39)

$$\{S_1(t), S_2(t),\ I_1(t),\ I_2(t),\ G_1(t),\ G_2(t),\ R_1(t),\ R_2(t)\}$$

are positive for all $t > 0$. □

Also, Let denote by S the total size of all individuals:

$$S(t) = S_1(t) + S_2(t) + I_1(t) + I_2(t) + G_1(t) + G_2(t) + R_1(t) + R_2(t)$$

Then, by adding all the model Equation (3.39), we have:

$$\frac{dS}{dt} = \beta_{11}\ (S_1 + G_1 + R_1) + \beta_{21}\ (S_2 + G_2 + R_2) - (\mu_2{}^S S_2 + \mu_2{}^I I_2 + \mu_2{}^G G_2 \ + \epsilon G_2 + \mu_2{}^R R_2$$

Let us denote $\mu = inf\{\mu_2^S, \mu_2^I, \mu_2^G + \epsilon, \mu_2^R\}$. By neglecting the fecundity rate of the young class and if $\beta = \beta_{21} \leq \mu$, we have:

$$\frac{dS}{dt} \leq \beta(S_2 + I_2 + G_2 + R_2) - -\mu(S_2 + I_2 + G_2 + R_2) \leq 0,$$

Then, we can conclude that the total size S is bounded, which implies the boundedness of the partial sizes $S_1(t), S_2(t),\ I_1(t),\ I_2(t),\ G_1(t),\ G_2(t),\ R_1(t),\ R_2(t)$.

3.6.3 Disease-Free (Eradication) equilibrium and stability of the endemic state

Setting the right hand side of the equations in the model Equation (3.39) to zero, i.e.,

$$\frac{dS_1}{dt} = \frac{dS_2}{dt} = \frac{dI_1}{dt} = \frac{dI_2}{dt} = \frac{dG_1}{dt} = \frac{dG_2}{dt} = \frac{dR_1}{dt} = \frac{dR_2}{dt} = 0$$

and supposing that all infectious class sizes are equal to zero which means that there is no disease (eradication) in the studied population, thus the disease-free equilibrium state is:

$$({S_1}^*,\ {S_2}^*, {I_1}^*, {I_2}^* {G_1}^*, {G_2}^*, {R_1}^*, {R_2}^*) = \left\{ \frac{\beta_{21} S_2}{a_1^S + \theta_1^G + \ \theta_1^R - \beta_{11}}, \frac{a_1^S S_1}{{\mu_2}^S + \theta_2^G + \ \theta_2^R}, 0, 0, 0, 0, 0, 0 \right\}$$

To show the stability of the endemic state, let us fix a set of values for the model parameters, where k is a scale parameter:

$$\varphi_1 \;=\; \varphi_2 \;=\frac{4k^2}{100},\; {\mu_2}^S=\epsilon_1 \;=\beta_{22} \;=0,\; {\mu_2}^I=\frac{499k}{100},\; {\mu_2}^G=\frac{49k}{100},\; {\mu_2}^R=\frac{2k}{5},\; \beta_{11} \;=\; \beta_{21} \;=\frac{k}{5},$$

$$\theta_1^G=\theta_1^R=\gamma_1=\gamma_2=0.1k,\; \theta_2^G=\theta_2^R=k,\; a_1^S=\; a_1^I=\; a_1^G=\; a_1^R=\frac{98k}{96},\; \epsilon_2 \;=\frac{4k}{5},\; \eta_1=\; \eta_2=0.2k$$

The two stationary points are labeled with * (resp. **) for the eradication (resp. endemic) state:

$$({S_1}^*,\; {S_2}^*, {I_1}^*, {I_2}^* {G_1}^*, {G_2}^*, {R_1}^*, {R_2}^*) = \left(\frac{\beta_{21}\; S_2}{a_1^S+\theta_1^G+\;\theta_1^R-\beta_{11}}, \frac{a_1^S S_1}{{\mu_2}^S+\theta_2^G+\;\theta_2^R}, 0,0,0,0,0,0\right)$$

and

$$({S_1}^{**},\; {S_2}^{**}, {I_1}^{**}, {I_2}^{**}, {G_1}^{**}, {G_2}^{**}, {R_1}^{**}, {R_2}^{**}) = (10,10,10,10,20,40,15,15)\,,\text{ if } k=1.$$

Let us show that the endemic state is locally stable. With the chosen parameter values and $k=1$, from the model Equation (3.39), we have:

$$\begin{aligned}
\frac{dS_1}{dt} &= 0.2G_1-1.02S_1+0.2R_1+0.2S_2+0.2G_2+0.2R_2-0.04I_1S_1-0.04I_2S_1\\
\frac{dS_2}{dt} &= 1.02S_1-2S_2-0.04I_1S_2-0.04I_2S_2\\
\frac{dI_1}{dt} &= 0.04I_1S_1+0.04I_2S_1-1.04I_1\\
\frac{dI_2}{dt} &= 0.04I_1S_2+0.04I_2S_2+1.02I_1-5.19I_2\\
\frac{dG_1}{dt} &= 0.2I_1+0.1S_1-1.12G_1\\
\frac{dG_2}{dt} &= 0.2I_2+1.02G_1+\;S_2-1.39G_2\\
\frac{dR_1}{dt} &= 0.1G_1+0.1S_1-1.02R_1\\
\frac{dR_2}{dt} &= 0.1G_2+S_2+1.02R_1-0.4R_2
\end{aligned} \tag{3.40}$$

By calculating the Jacobian matrix M of the system of Equation (3.40) at the endemic state, where I is the identity matrix and finding the roots of its characteristic polynomial P_M, we have for the second stationary point the expression as follows:

$$M-\lambda I=\begin{pmatrix}
-1.02-\lambda & 0.2 & -0.4 & -0.4 & 0.2 & 0.2 & 0.2 & 0.2\\
1.02 & -2-\lambda & -0.4 & -0.4 & 0 & 0 & 0 & 0\\
0.8 & 0 & -1.04-\lambda & 0 & 0 & 0 & 0 & 0\\
0 & 0.8 & 1.02 & -5.19-\lambda & 0 & 0 & 0 & 0\\
0.1 & 0 & 0.2 & 0 & -1.12-\lambda & 0 & 0 & 0\\
0 & 1 & 0 & 0.2 & 1.02 & -1.39-\lambda & 0 & 0\\
0.1 & 0 & 0 & 0 & 0.1 & 0 & -1.02-\lambda & 0\\
0 & 1 & 0 & 0 & 0 & 0.1 & 1.02 & -0.4-\lambda
\end{pmatrix}$$

The roots of the characteristic polynomial of M, P_M, satisfy:

$$\begin{aligned} P_M(\lambda) &= \det(M-\lambda I) = (\lambda+0.706923)(\lambda+5.13103)\left(\lambda^2+0.973528\lambda+0.34052\right) \\ &\quad \left(\lambda^2+2.0371\lambda+1.03861\right)\left(\lambda^2+3.69142\lambda\ +3.53083\right) \\ &= 0 \end{aligned}$$

The real parts of the eigenvalues of the matrix M are all negative, equal to:

$$-0706923, -5.13103, -0.486764, -1.01855, -1.84571$$

Hence, with the set of chosen parameter values, the stability of the endemic state is proved.

3.6.4 Existence and unicity of the solution

We want to establish existence and unicity of the solution for the model Equation (3.39). Let us denote the second member of Equation (3.39) by $\mathcal{H} = (\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3, \mathcal{H}_4, \mathcal{H}_5, \mathcal{H}_6, \mathcal{H}_7, \mathcal{H}_8)$ and the state vector by $Z = (S_1, S_2,\ I_1,\ I_2,\ G_1,\ G_2,\ R_1,\ R_2)$, with $Z(0) =\ Z_0,\ Z(t,\epsilon) = \mathcal{H}(t,\epsilon)\, Z(t)$.
Let us define the Volterra integral equation of the second kind which is given as:

$$S(t) =\ S(0) + \int_0^t \mathcal{H}(t,\epsilon)\, S(\epsilon)\, d\epsilon$$

where $\mathcal{H}(t,\epsilon)$ is a kernel and $S(t)$ is the function to be solved.
The Volterra integral equation formulation of the model Equation (3.39) is the following:

$$\begin{aligned} S_1(t) &= S_1(0) + \int_0^t \mathcal{H}_1(t,\epsilon)\, S_1(\epsilon)\, d\epsilon \\ S_2(t) &= S_2(0) + \int_0^t \mathcal{H}_2(t,\epsilon)\, S_2(\epsilon)\, d\epsilon \\ I_1(t) &= I_1(0) + \int_0^t \mathcal{H}_3(t,\epsilon)\, I_1(\epsilon)\, d\epsilon \\ I_2(t) &= I_2(0) + \int_0^t \mathcal{H}_4(t,\epsilon)\, I_2(\epsilon)\, d\epsilon \\ G_1(t) &= G_1(0) + \int_0^t \mathcal{H}_5(t,\epsilon)\, G_1(\epsilon)\, d\epsilon \\ G_2(t) &= G_2(0) + \int_0^t \mathcal{H}_6(t,\epsilon)\, G_2(\epsilon)\, d\epsilon \\ R_1(t) &= R_1(0) + \int_0^t \mathcal{H}_7(t,\epsilon)\, R_1(\epsilon)\, d\epsilon \\ R_2(t) &= R_2(0) + \int_0^t \mathcal{H}_8(t,\epsilon)\, R_2(\epsilon)\, d\epsilon \end{aligned} \tag{3.41}$$

Let $\mathcal{S}_1$ be the solution for S_1 and kernels $\mathcal{H}_i,\ i = 1, 2, \ldots, 8$ satisfy Lipschitz conditions:

$$\sup_{0<t\leq 1}\|S_1\| \leq c_1, \sup_{0<t\leq 1}\|S_2\| \leq c_2, \sup_{0<t\leq 1}\|I_1\| \leq c_3,\ \sup_{0<t\leq 1}\|I_2\| c_4,$$

$$\sup_{0<t\leq 1}\|G_1\| \leq c_5, \sup_{0<t\leq 1}\|G_2\| \leq c_6,\ \sup_{0<t\leq 1}\|R_1\| \leq c_7,\ \sup_{0<t\leq 1}\|R_2\| \leq c_8$$

and $c_i > 0,$ for $i = 1,\ 2, \ldots, 8$. Then, following inequalities hold, using triangle inequality and properties of the $\mathcal{H}_i'$s norm:

$$\begin{aligned}
\|\mathcal{H}_1(\mathrm{S}_1) - \mathcal{H}_1(\mathcal{S}_1)\| &\leq \|\left[\beta_{11} - \left(a_1^S + \theta_1^G + \ \theta_1^R + \ \varphi_1(I_1 + I_2)\right)\right] S_1 \\
&\quad - \left[\beta_{11} - (a_1^S + \theta_1^G + \ \theta_1^R + \ \varphi_1(I_1 + I_2))\right] \mathcal{S}_1\| \\
&\leq \|\beta_{11} - (a_1^S + \theta_1^G + \ \theta_1^R + \ \varphi_1(I_1 + I_2))\|\|S_1 - \mathcal{S}_1\| \\
&\leq \|\beta_{11} - (a_1^S + \theta_1^G + \ \theta_1^R + \ \varphi_1 c_3 + \varphi_1 c_4)\|\|S_1 - \mathcal{S}_1\| \\
&= \partial_1\|S_1 - \mathcal{S}_1\|
\end{aligned}$$

where $\xi = \beta_{11}\ (G_1 + R_1) + \beta_{21}(S_2 + G_2 + R_2), \partial_1 = \beta_{11} - a_1^S - \theta_1^G - \ \theta_1^R - \ \varphi_1 c_3 - \varphi_1 c_4$. Therefore, $\mathcal{H}_1$ satisfies the Lipschitz conditions. We can show in the same way that other functions $\mathcal{H}_i,\ i = 2, \ldots, 8$ in the model Equation (3.39) satisfy the Lipschitz conditions as follows:

$$\|\mathcal{H}_2(S_2) - \mathcal{H}_2(\mathcal{S}_2) \leq \|\ \partial_2\|S_2 - \mathcal{S}_2\|$$

$$\|\mathcal{H}_3(I_1) - \mathcal{H}_3(\mathcal{L}_1) \leq \ \|\partial_3\|I_1 - \mathcal{L}_1\|$$

$$\|\mathcal{H}_4(I_2) - \mathcal{H}_4(\mathcal{L}_2) \leq \|\ \partial_4\|I_2 - \mathcal{L}_2\|$$

$$\|\mathcal{H}_5(G_1) - \mathcal{H}_5(\mathcal{G}_1) \leq \|\ \partial_5\|G_1 - \mathcal{G}_1\|$$

$$\|\mathcal{H}_6(G_2) - \mathcal{H}_6(\mathcal{G}_2) \leq \|\ \partial_6\|G_2 - \mathcal{G}_2\|$$

$$\|\mathcal{H}_7(R_1) - \mathcal{H}_7(\Re_1) \leq \|\ \partial_7\|R_1 - \Re_1\|$$

$$\|\mathcal{H}_8(R_2) - \mathcal{H}_8(\Re_2) \leq \|\ \partial_8\|R_2 - \Re_2\|$$

Let us now consider the following Neumann series:

$$S_{1m}(t) = S_1(0) + \int_0^t \mathcal{H}_1(t,\epsilon)\, S_{1m-1}(\epsilon)\, d\epsilon$$

$$S_{2m}(t) = S_2(0) + \int_0^t \mathcal{H}_2(t,\epsilon)\, S_{2m-1}(\epsilon)\, d\epsilon$$

$$I_{1m}(t) = I_1(0) + \int_0^t \mathcal{H}_3(t,\epsilon)\, I_{1m-1}(\epsilon)\, d\epsilon$$

$$I_{2m}(t) = I_2(0) + \int_0^t \mathcal{H}_4(t,\epsilon)\, I_{2m-1}(\epsilon)\, d\epsilon$$

$$G_{1m}(t) = G_1(0) + \int_0^t \mathcal{H}_5(t,\epsilon)\, G_{1m-1}(\epsilon)\, d\epsilon$$

$$G_{2m}(t) = G_2(0) + \int_0^t \mathcal{H}_6(t,\epsilon)\, G_{2m-1}(\epsilon)\, d\epsilon$$

$$R_{1m}(t) = R_1(0) + \int_0^t \mathcal{H}_7(t,\epsilon)\, R_{1m-1}(\epsilon)\, d\epsilon$$

$$R_{2m}(t) = R_2(0) + \int_0^t \mathcal{H}_8(t,\epsilon)\, R_{2m-1}(\epsilon)\, d\epsilon$$

These Neumann series are convergent due to the Lipschtizian character of $\mathcal{H}'$s, then:

$$\|S_{1m+1} - S_{1m}\| \leq \int_0^t \|\mathcal{H}_1(t,\epsilon)\, S_{1m}(\epsilon) - \mathcal{H}_1(t,\epsilon), S_{1m-1}(\epsilon)\, d\epsilon\|$$

$$\leq \int_0^t \|S_{1m}(t) - S_{1m-1}(t)\| d\epsilon \ \leq \partial_1 \|S_{1m}(t) - S_{1m-1}(t)\|_\infty$$

Other equations are given as follows:

$$\|S_{2m+1} - S_{2m}\| \ \leq \partial_2 \|S_{2m}(t) - S_{2m-1}(t)\|_\infty$$

$$\|I_{1m+1} - I_{1m}\| \ \leq \partial_3 \|I_{1m}(t) - I_{1m-1}(t)\|_\infty$$

$$\|I_{2m+1} - I_{2m}\| \ \leq \partial_4 \|I_{2m}(\mathrm{t}) - I_{2m-1}(t)\|_\infty$$

$$\|G_{1m+1} - G_{1m}\| \ \leq \partial_5 \|G_{1m}(t) - G_{1m-1}(t)\|_\infty$$

$$\|G_{2m+1} - G_{2m}\| \ \leq \partial_6 \|G_{2m}(t) - G_{2m-1}(t)\|_\infty$$

$$\|R_{1m+1} - R_{1m}\| \ \leq \partial_7 \|R_{1m}(t) - R_{1m-1}(t)\|_\infty$$

$$\|R_{2m+1} - R_{2m}\| \ \leq \partial_8 \|R_{2m}(t) - R_{2m-1}(t)\|_\infty$$

The above inequalities prove the existence of the function $\mathcal{H}$. We now show the uniqueness of the solution by assuming that the kernels $\mathcal{H}_i,\ i = 1, 2, \ldots, 8$ are separable, i.e., $\mathcal{H}_1(t,\epsilon) = \varnothing(\mathrm{t})\,\varsigma(\epsilon)$. By denoting:

$$\zeta(\mathrm{t}) = \int_0^t S_1(\epsilon)\,\varsigma(\epsilon)\, d\epsilon$$

then, because $S_1(t) = S_1(0) + \int_0^t \mathcal{H}_1(t,\epsilon)\, S_1(\epsilon)\, d\epsilon = S_1(0) + \int_0^t \varnothing(t)\, \varsigma(\epsilon)\, S_1(\epsilon)\, d\epsilon$,
we have:

$$\zeta\,\prime(t) = \varnothing(t)\,\varsigma(t)\,\zeta(t) + \ S_1(0)\,\varsigma(t)$$

If $\varsigma(0) = 0$, the solution $\zeta(t)$ follows $\zeta(t) = \ e^{\int_0^t \varnothing(\epsilon)\varsigma(\epsilon)d\epsilon}$ and from the definition of $\zeta(t)$, the solution of $S_1(t)$ is given as:

$$S_1(t) = S_1(0) + \varnothing(t)\int_0^t S_1(\epsilon)\,\varsigma(\epsilon)\,d\epsilon = \ S_1(0) + \varnothing(t)\,\zeta(t) = S_1(0) + \varnothing(t)\, e^{\int_0^t \varnothing(\epsilon)\varsigma(\epsilon)d\epsilon}$$

Hence, by the unicity of the solution $\zeta(t)$, there exists just only one continuous solution for $S_1(t)$. Following the same approach, we can obtain a unique solution for the remaining equations of the system in Equation (3.41).

3.6.5 Basic reproduction number

It is possible to apply the idea of a next-generation matrix by linearizing the model Equation (3.39) near the endemic state, for the infectious part of the system in Equation (3.39), and then obtain:

$$\begin{aligned}
\frac{dI_1}{dt} &= \varphi_1(I_1 + I_2)\, {S_1}^{**} - (a_1^I + \eta_1) I_1, \\
\frac{dI_2}{dt} &= \varphi_2(I_1 + I_2)\, {S_2}^{**} + a_1^I I_1 - \eta_2 I_2 \ \ {\mu_2}^I I_2, \\
\frac{dG_1}{dt} &= \ \eta_1 I_1 + \ \theta_1^G S_1^{**} - \left(a_1^G + \gamma_1 + \epsilon_1\right) G_1, \\
\frac{dG_2}{dt} &= \ \eta_2 I_2 + \ a_1^G G_1 + \ \theta_2^G {S_2}^{**} - \left(\gamma_2 + \ {\mu_2}^G + \epsilon_2 + \beta_{22}\right) G_2
\end{aligned} \tag{3.42}$$

By summing Equation (3.42) and by denoting I as the size of all infectious, we have:

$$I(t) = I_1(t) + I_2(t) + G_1(t) + G_2(t),$$

and

$$\frac{dI}{dt} = \varphi_1(I_1 + I_2)\, {S_1}^{**} + \varphi_2(I_1 + I_2)\, {S_2}^{**} \ - {\mu_2}^I I_2 + \ \theta_1^G {S_1}^{**} - (\gamma_1 + \epsilon_1) G_1 \ - (\gamma_2 + \epsilon_2 + \beta_{22}) G_2$$

The matrix J of the linearized system near a state (S_1, S_2) is:

$$J = \begin{bmatrix}
\varphi_1 S_1 - a_1^I - \eta_1 & \varphi_1 S_1 & 0 & 0 \\
\varphi_2 S_2 + a_1^I & \varphi_2 S_2 - \eta_2 - \ {\mu_2}^I & 0 & 0 \\
\eta_1 & 0 & -a_1^G - \gamma_1 - \epsilon_1 & 0 \\
0 & \eta_2 & a_1^G & -\gamma_2 - \ {\mu_2}^G - \epsilon_2 - \beta_{22}
\end{bmatrix}$$

The corresponding characteristic polynomial $P_J(\lambda)$ is equal to:

$$\begin{aligned}
&(\varphi_1 S_1 - a_1^I - \eta_1 - \ \lambda)(\varphi_2 S_2 - \eta_2 - \ {\mu_2}^I - \ \lambda)\ \left(-a_1^G - \gamma_1 - \epsilon_1 - \ \lambda\right)\ \left(-\mu_2^G - \gamma_2 - \epsilon_2 - \beta_{22} - \ \lambda\right) \\
&\quad - \varphi_1 t S_1 \left(\varphi_2 S_2 + a_1^I\right)\left(-a_1^G - \gamma_1 - \epsilon_1 - \ \lambda\right)\left(-\mu_2^G - \gamma_2 - \epsilon_2 - \beta_{22} - \lambda\right) = P_J(\lambda)
\end{aligned}$$

The positive eigenvalues of J are roots of the following polynomial:

$$\lambda^2-\left(\varphi_1 S_1 - a_1^I - \eta_1 + \varphi_2 S_2 - \eta_2 - \ {\mu_2}^I\right)\lambda - \ \varphi_1 S_1\left(\varphi_2 S_2 + a_1^I\right) + (\varphi_1 S_1 - a_1^I - \eta_1)(\varphi_2 S_2 - \eta_2 - \ {\mu_2}^I)$$

These roots are equal to: $B \pm (B^2 - C)^{1/2}$, where B and C are equal to:

$$B = \frac{\varphi_1 S_1 - a_1^I - \eta_1 + \varphi_2 S_2 - \eta_2 - \ {\mu_2}^I}{2}$$

$$C = - \ \varphi_1 S_1\left(\varphi_2 S_2 + a_1^I\right) + (\varphi_1 S_1 - a_1^I - \eta_1)(\varphi_2 S_2 - \eta_2 - \ {\mu_2}^I)$$

Then, we have:

$$\begin{aligned} B + (B^2 - C)^{1/2} \ = \ & \frac{\varphi_1 S_1 - a_1^I - \eta_1 + \varphi_2 S_2 - \eta_2 - \ {\mu_2}^I}{2} \\ & + \Big[\frac{(\varphi_1 S_1 - a_1^I - \eta_1 + \varphi_2 S_2 - \eta_2 - \ {\mu_2}^I)^2}{4} \\ & + \ \varphi_1 S_1\left(\varphi_2 S_2 + a_1^I\right) - \left(\varphi_1 S_1 - a_1^I - \eta_1\right)\left(\varphi_2 S_2 - \eta_2 - \ {\mu_2}^I\right) \Big]^{1/2} \end{aligned}$$

Hence, near unstable endemic state the positive dominant eigenvalue Λ is equal to:

$$\begin{aligned} \Lambda \ = \ & \frac{\varphi_1 S_1 - a_1^I - \eta_1 + \varphi_2 S_2 - \eta_2 - \ \mu_2^I}{2} \\ & + \left[\frac{(\varphi_1 S_1 - a_1^I - \eta_1 - \varphi_2 S_2 + \eta_2 + \ \mu_2^I)^2}{4} + \ \varphi_1 S_1\left(\varphi_2 S_2 + a_1^I\right) \right]^{1/2} \end{aligned}$$

Therefore, the basic reproduction number R_0 equal to Λ near the endemic stationary state depends mainly on the infection rates φ_1 and φ_2, when the sizes S_1 and S_2 are sufficiently important. If after a change of parameter values, the endemic state becomes unstable and R_0 is becoming more than 1, then an epidemic wave starts.

3.6.6 Numerical simulation of the model: examples of the COVID-19 outbreak in some countries

The numerical simulation of the model developed is done using data coming from Kuwait, France and Cameroon. First, we provide some explanations for the data used for the simulations, including how we assumed some of the parameters, calculated others, and selected some from the literature cited in this article. In order to determine the susceptible classes sizes, we used Wikipedia data on the three studied countries to obtain their total population size. We then calculated the ratio between the young and elderly from the data presented in Figure 3.38, Figure 3.40 and Figure 3.39 and used this ratio to determine the sizes value for the two susceptible classes at the exponential phase considered (for Kuwait 28 December 2020, France 30 October 2021, and Cameroon 19 September 2021), because the Ross and Kermack-McKendrick model is only suited for the exponential growth phase of an epidemic wave. Because the progression rates to the classes reversed recovery and fully recovered in Figure 3.41 are not zero, we assumed that they should have some populations at the start of the wave, even if they were small. The transmission rate for Cameroon was chosen at the

start of the second wave in January 2021 from [119, 151], for France at the start of the fifth wave in December 2021 [217], and for Kuwait at the start of the fourth wave in December 2021 [214]. The values of specific death rates due to disease and natural death rates for Kuwait and France were taken from [160, 214, 217]. For Cameroon, the natural death rate was taken from [218], while the specific death rate due to disease was calculated from the cumulated deaths number due to the disease in two years divided by the cumulated infectious number in these two years [218]. The vaccination rate was chosen from [126], while the loss of resistance was chosen from [214]. Other parameters were assumed.

Numerical simulation based on COVID-19 outbreak in Kuwait

We have chosen the following set of parameter values corresponding to the COVID-19 outbreak at the start of the fourth wave in December 2021 in Kuwait:

$$\begin{aligned}
&\varphi_1 = 1.7, \varphi_2 = 0.9, {\mu_2}^S = 0.003, \epsilon_1 = 0.28, \beta_{22} = 0, \ {\mu_2}^I = 0.0025, \\
&{\mu_2}^G = 0.002, \ {\mu_2}^R = 0.0021 \ , \ \beta_{11} = 2.1 \ , \beta_{21} = 2.3 \ , \ \theta_1^G = 0.765, \theta_1^R = 0.678, \\
&\gamma_1 = 0.62, \gamma_2 = 0.74, \ \theta_2^G = 0.45 \ , \theta_2^R = 0.33, \ a_1^S = 0.7, a_1^I = 0.54, a_1^G = 0.65, \\
&a_1^R = 0.8, \ \epsilon_2 = 0.19 \ , \eta_1 = 0.3 \text{ and } \eta_2 = 0.38,
\end{aligned}$$

with initial values for $S_1 = 4413099, S_2 = 51422, I_1 = 65, I_2 = 194, G_1 = 20, G_2 = 17, R_1 = 30$ and $R_2 = 73$.

We present the visualization results for the simulated values in Figure 3.42.

Numerical simulation based on COVID-19 outbreak in France

We have chosen the following set of parameter values corresponding to the COVID-19 outbreak at the start of the fifth wave in December 2021 in France:

$$\begin{aligned}
&\varphi_1 = 1.2, \varphi_2 = 0.9, {\mu_2}^S = 0.009, \epsilon_1 = 0.28, \beta_{22} = 0, \ {\mu_2}^I = 0.0025, \\
&{\mu_2}^G = 0.002, \ {\mu_2}^R = 0.0021 \ , \ \beta_{11} = 1.9, \beta_{21} = 2.2, \ \theta_1^G = 0.735, \theta_1^R = 0.678, \\
&\gamma_1 = 0.62, \gamma_2 = 0.74, \ \theta_2^G = 0.45, \theta_2^R = 0.33, \ a_1^S = 0.7, a_1^I = 0.54, a_1^G = 0.65, \\
&a_1^R = 0.8, \ \epsilon_2 = 0.19, \eta_1 = 0.3 \text{ and } \eta_2 = 0.38,
\end{aligned}$$

with initial values for $S_1 = 53372880, S_2 = 14017120, I_1 = 3912, I_2 = 1757, G_1 = 200, G_2 = 170, R_1 = 300$ and $R_2 = 730$.

We present the visualization results for the simulated values in Figure 3.43.

Numerical simulation based on COVID-19 outbreak in Cameroon

We choose the following set of parameter values corresponding to the COVID-19 outbreak at the start of the second wave in January 2021 in Cameroon:

$$\begin{aligned}
&\varphi_1 = 1.3, \varphi_2 = 0.9, {\mu_2}^S = 0.009, \epsilon_1 = 0.28, \beta_{22} = 0, \ {\mu_2}^I = 0.0025, \\
&{\mu_2}^G = 0.002, \ {\mu_2}^R = 0.0021 \ , \ \beta_{11} = 4.5, \beta_{21} = 4.1, \ \theta_1^G = 0.024, \theta_1^R = 0.678, \\
&\gamma_1 = 0.62, \gamma_2 = 0.74, \ \theta_2^G = 0.45, \ \theta_2^R = 0.33, \ a_1^S = 0.7, a_1^I = 0.54, a_1^G = 0.65, \\
&a_1^R = 0.8, \ \epsilon_2 = 0.19, \eta_1 = 0.3 \text{ and } \eta_2 = 0.38,
\end{aligned}$$

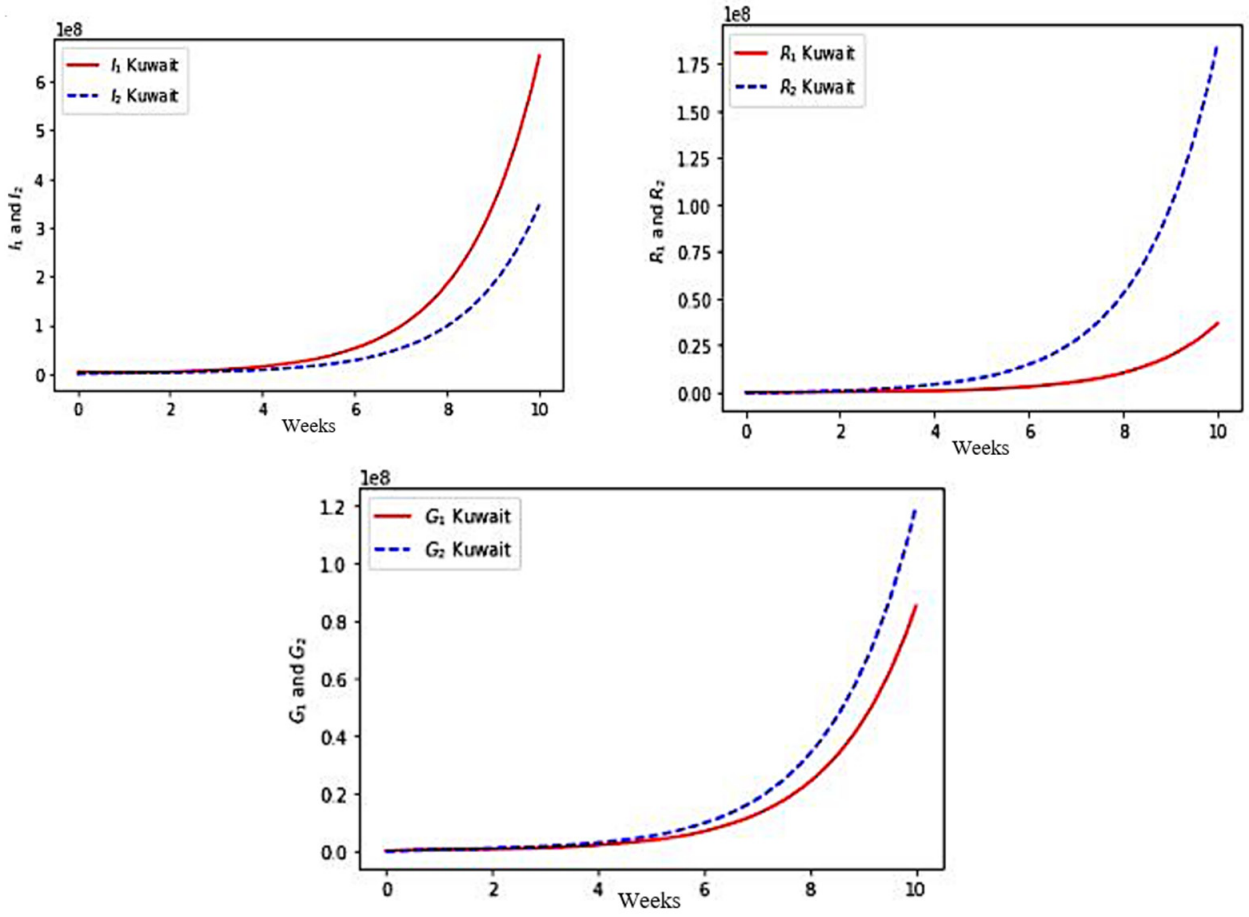


Figure 3.42: Numerical simulation of the variables I_1 and I_2, R_1 and R_2, G_1 and G_2 for Kuwait.

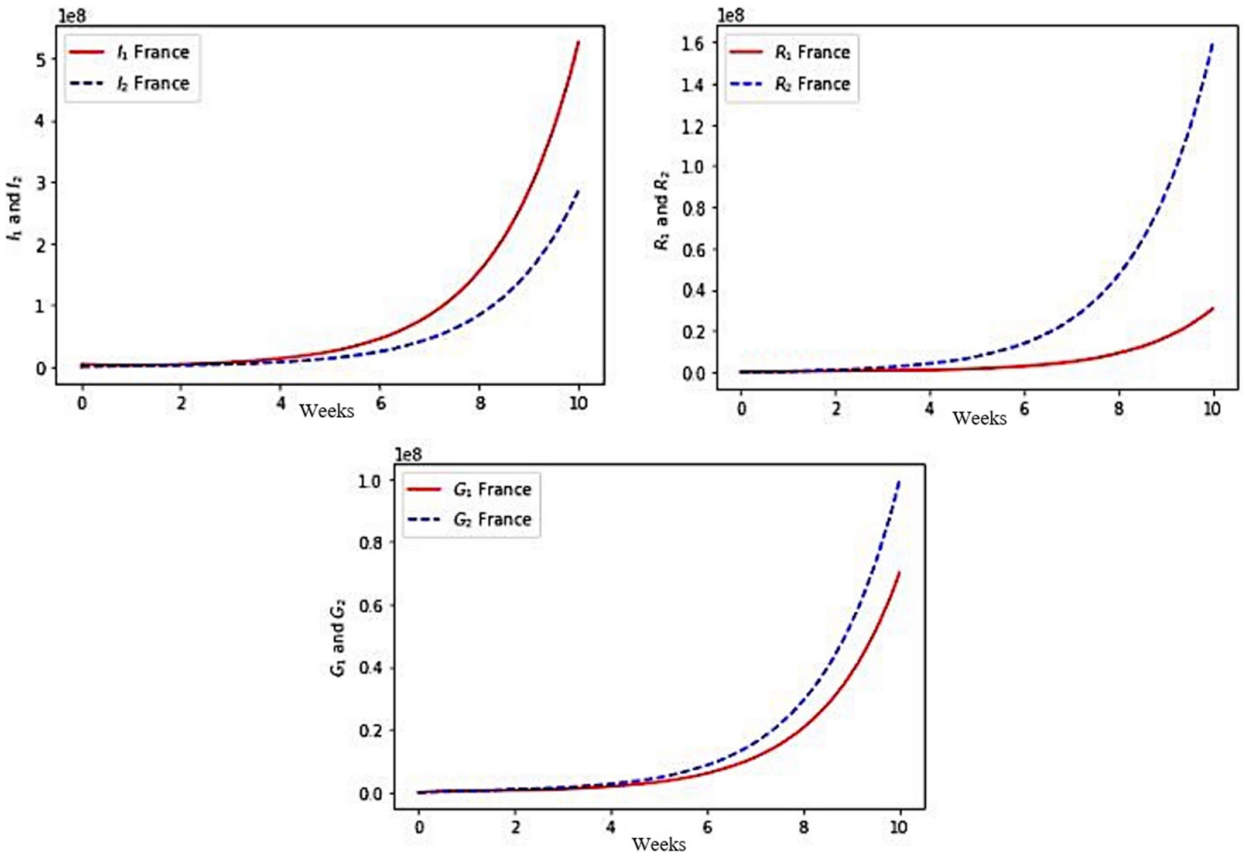


Figure 3.43: Numerical simulation of the variables I_1 and I_2, R_1 and R_2, G_1 and G_2 for France.

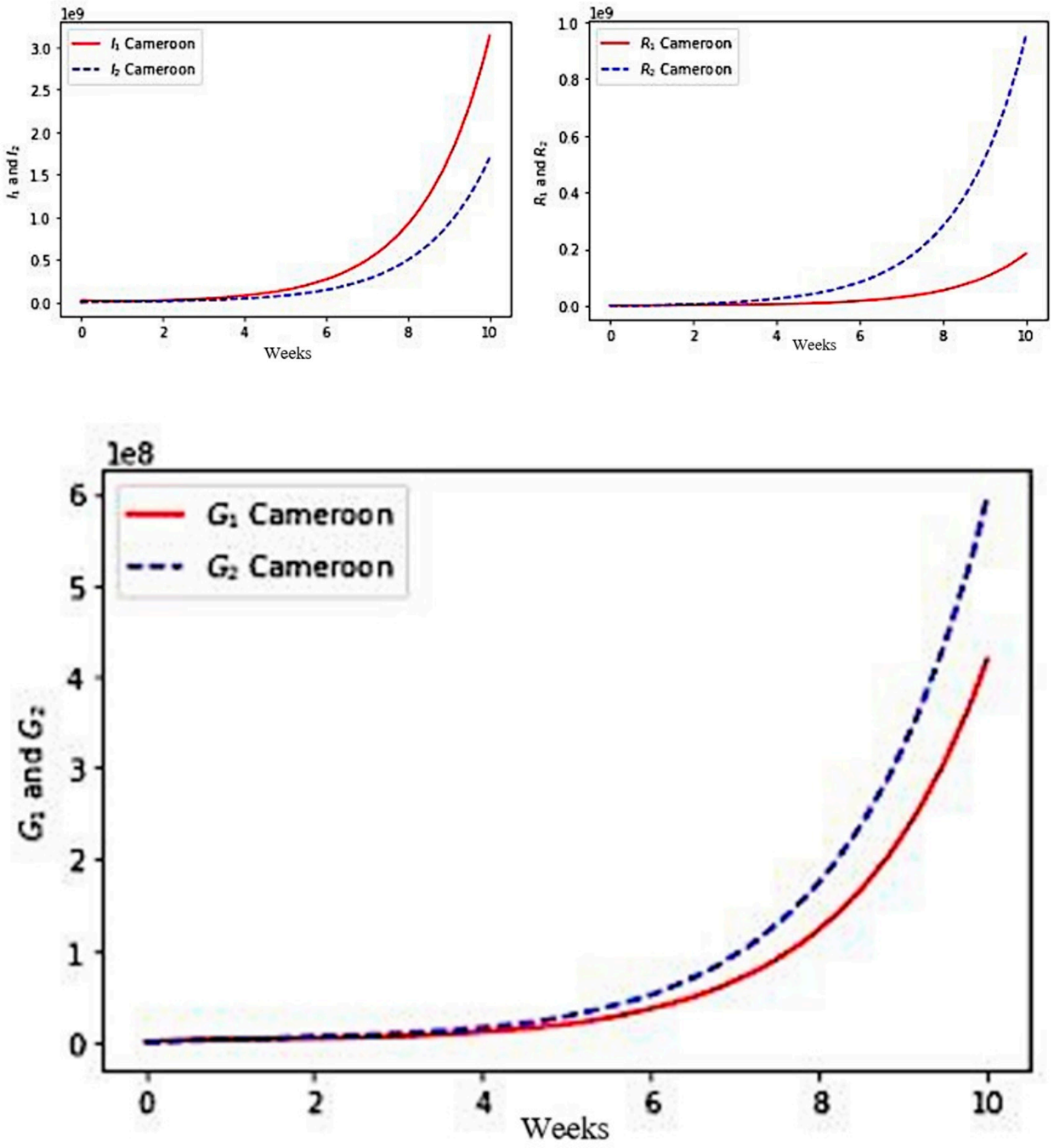


Figure 3.43: Numerical simulation of the variables I_1 and I_2, R_1 and R_2, G_1, G_2 for Cameroon.

with initial values for $S_1 = 25828973, S_2 = 721027, I_1 = 151, I_2 = 21, G_1 = 75, G_2 = 33, R_1 = 56$ and $R_2 = 64$.

We present the visualization results for the simulated values in Figure 3.43.
The results for the three considered countries present some similarities but also some differences that we will discuss in the following.
First, in each case, the exponential growth of the infectious $I_1 + I_2$ and completely recovering $R_1 + R_2$ populations sizes correspond roughly to the data given in [119]. The model simulations give more, i.e., allows to see the part brought by each age class to the global

growth.
Second, for Kuwait, Figure 3.42 shows a faster growth of infectious (I) in the young class ($\leq$65 years) than in the older class ($>$65 years), which is also the case for France (Figure 3.43) and for Cameroon (Figure 3.43). On the other hand, this phenomenon is reversed in the three countries, with regard to the growth of populations immunized in a transient manner (G) and in a lasting manner (R). The phenomenon is more marked for Kuwait than for France, France itself having a more marked difference than for Cameroon. This is partly explained by the better vaccination rate of Kuwait than that of France and Cameroon in the young class, the effectiveness of the vaccination (which decreases in the older class) having been assumed to be equal for the three countries.
Taking age into account in modeling the COVID-19 pandemic makes it possible to simulate the differential dynamic behavior of the growth of infectious and immune populations, young and old, in order, for example, to adjust the vaccine policy according to the age.

3.7 Conclusion

Concerning contagious diseases, public health physicians are constantly faced with three challenges. The first concerns the estimation of the basic reproduction number R_0. The systematic use of R_0 simplifies the decision-making process by policymakers, advised by public health authorities, but it is too much of a caricature to account for the biology behind the viral spread. We have observed in the COVID-19 outbreak that it was non-constant during an epidemic wave due to exogenous and endogenous factors influencing both the duration of the contagiousness period and the daily transmission rate during this phase [219, 220, 221]. Then, the second challenge concerns the estimation of the mean duration of the infectious period for infected patients. As for the transmission rate, realistic assumptions made it possible to obtain an upper limit to this duration [137], mainly due to the lack of viral load data in large patient cohorts (see Figure 3.2 from [116, 117, 118]), in order to better guide the individual quarantine measures decided by the authorities in charge of public health. This upper bound also makes it possible to obtain a lower bound for the percentage of unreported infected patients, which gives an idea of the quality of the census of cases of infected patients, which is the second challenge facing specialists of contagious diseases. The third challenge is the estimation of the daily reproduction number over the contagiousness period, which was one of the issue we tackle in this chapter.
Eventually, our approach using marginal daily reproduction numbers involving a certain level of noise in the dynamics of new daily infected cases defines a stochastic framework which describes phenomenologically the exponential phase as our results show for countries such as France, Russia, Sweden, etc. This stochastic modeling allows a better understanding of the role of the contagiousness period length and of the heterogeneity (e.g., the U-shape) of its daily reproduction number distribution in the COVID-19 outbreak dynamics. On the medical level, the important message about the U-shape is that COVID-19 is similar to other viral diseases, such as influenza, with two successive reactions from the two immune defense barriers, innate cellular immunity first, which is not sufficient if symptoms persist, then adaptive immunity (cellular and humoral), which results in a transient decrease in contagiousness between the two phases. The medical recommendations are, in this case, never to take a transient improvement for a permanent disappearance of the symptoms. One could indeed, for a public health use, be satisfied after estimating the sum of the R_j's, that is to say, R_0 or the effective R_e. For an individual health use, it is important to

know the existence of a minimum of the R_j's, which generally corresponds to a temporary clinical improvement, after the partial success of the innate immune defenses. This makes it possible to prevent the patient from continuing to respect absolute isolation and therapeutic measures, even if a transient improvement occurs; otherwise, they risk, as in the flu, a bacterial pulmonary super-infection (a frequent cause of death in the case of COVID-19). On the theoretical level, the interest of the proposed methods is its generic character: it can be applied to all contagious diseases, within the very general framework of Equation (3.6), which makes no assumption about the spatial heterogeneity or the longitudinal constancy of the daily reproduction numbers. The deconvolution of Equation (3.6) poses a new theoretical problem when it is offered in this context.

The current context of the pandemic closely monitored by the WHO, a contagious social pandemic and a contagious infectious pandemic, COVID-19, shows that this problem is now highly relevant.

To be able to be both retro-predictive on past pandemics (like the great plague of 1348 in Europe) or anticipatory on present and future pandemics, it is necessary to be able, from observed data (which are daily in the case of the COVID-19 [119] and annual in the case of obesity [214]), to estimate the parameters of the model, then simulate it and interpret the scenarios of the simulations within the framework of concrete public health measures (change in eating behaviours, preventive preparation of the system immunity, implementation of mitigation measures, vaccination, preventive education etc.). This part is the most difficult and is only effective within the framework of multidisciplinary teams, with very good mathematicians of dynamical systems and very good nutritionists and infectiologists. Solid knowledge in the genetics of infectious agents and in the monitoring of environmental conditions is necessary because:

- the dynamic phenomena of propagation present isolated or periodic waves of the new cases, with bifurcations (in the deterministic case) or phase transitions (in the stochastic case) often difficult to study theoretically and to simulate numerically [222, 223, 224]

- the changes in the infectious genius of the agents at the origin of pandemics need a new knowledge about mechanisms often partly unknown, because linked to uneasily predictable and often hypothetical [139, 225] modifications in their genome (such as point mutations, deletions and insertions of more or less long sequences, re-combinations with other genomes, passages in the host nuclear genome, etc.).

Furthermore, we have been able to use a continuous approach in estimating the maximal reproduction number in COVID-19 outbreak and by extension estimating the transmission rate, a parameter which helps to know how a disease is spread by identifying the point of inflexion during the exponential phase of the pandemic wave, we have then deduced the value of the basic reproduction rate at this point of inflexion while the method was validated by linking the correlation of these epidemiology parameters with some socio-demo-economic parameters.

Finally, we used a matrix population growth modeling on countries where age data were available for this study and also propose some panorama on cells targeted by SARS-CoV-2 and cell lifespan loss and also on how the method can be extended to ageing in cell populations using discrete and continuous approach. The model developed in this study differs from many previously constructed agent-based and dynamical system models of population dynamics, particularly when considering viral infection. Its greatest utility is that it can be used to highlight specific characteristics of the life span caused by viral infection, which have a significant impact on population growth. As a result, the findings of this study may be

useful in developing management programs to reduce the potential for COVID-19 epidemics, particularly when comorbidities and ageing are considered. We used data from Cameroon to investigate the role of comorbidities in age-dependent COVID-19 modeling, while data from France were used to assess the proportion of parameter sensitivity at the sub-population level. Data from Ireland were used to examine the evolution of the pandemic in young people because this country has seen a sharp increase in cases among children and teenagers as at the time of this study. Age dependent modeling is important to better understand the dynamics of viral diseases, and we have been able to present another perspective to this research direction using a discrete approach, matrix algebra and an ODE compartment model. It is a known fact that due to a new variant of the disease, there is an increase in infection rate in different age groups. Our work was able to provide a link between the dominant eigenvalue which represents the exponential growth parameter and the COVID-19 affected sub-population and also the differential dynamic behavior of the growth of infectious and immune populations between young and old.

Chapter 4

Statistical Modeling of COVID-19 Outbreak

In epidemiology, statistical modeling helps to understand the mechanisms that influence the spread of infectious diseases, and it suggests prevention and control strategies. It can be used to comprehend how a virus spreads across a large region or country hence, this chapter is devoted to large scale analysis and application of different statistical tools such as machine learning and deep learning to model the COVID-19 pandemics so as to predict and forecast the evolution of the disease at the population level. COVID-19 cases prediction models are currently divided into three categories: theoretical models, single artificial intelligence models, and decomposition integration models. The majority of statistical methods are used in the case study of COVID-19 to make prediction of the pandemic evolution at different phases. We present in this chapter some useful quantitative statistical tools for analysing epidemiological data.

Section 4.1 examines the application of ARIMA, regression models, exponential smoothing model and prophet model on several time series. The slope of the regression line is calculated using an exponential model, while the initial auto-correlation slope of new cases were forecasted using an ARIMA model for the two waves observed at the time of this study. The best statistical model to describe the daily count of new cases and deaths due to COVID-19 infections in Kuwait is presented by testing the model's performance, accuracy, and validation, as a result, the best model to estimate the number of new cases as well as a model for COVID-19 infection-related mortality were utilized. The findings indicate that the COVID-19 outbreak in Kuwait is divided into two stages i.e., first wave and transition to second wave. In Section 4.2, the application of deep learning to COVID-19 pandemic time series data has been done for the United States (USA), United Kingdom (UK), India, Russia, Brazil, France, and Turkey, which are countries where the pandemic is more prevalent globally. Multi-layer Perceptron (MLP), Long Short-term Memory (LSTM), Gated Recurrent Unit (GRU), Convolution Neural Network (CNN), and Deep Neural Network (DNN) methods were used on time series data from the beginning of the pandemic until September 3, 2021 to predict daily new cases and daily deaths at different waves of the pandemic in countries considered, with MLP forecasted until December 2021 due to the model performance. The performance of the methods was accessed using Root Mean Square Error (RMSE) and relative Root Mean Square Error (rRMSE). The results have depicted the impact of COVID-19 management across the countries studied. In Section 4.3 spectral analysis method is used to convert time (days) for COVID-19 pandemic time series data from United States (USA), United Kingdom

(UK), India, Russia, Brazil, France, and Turkey to analyze the frequency peaks and periodicity of the time series data. It was discovered that the pandemic's peaks detection was central in the analysis and short-term forecasting. In Section 4.4 the modelling of the COVID-19 pandemic against some socio-economic parameters is presented by examining the time series of infection spread throughout the first and second waves in all countries (OECD, developing and developed countries namely African countries), as well as their socio-economic statistics. The proof of a mathematical and statistical relationship between Theil and Gini indices is presented, and then multiple machine learning and deep learning approaches investigate the association between epidemiological data and socioeconomic variables. Also the relationship between epidemiologic and socio-economic variables such as CHE/GDP data is studied thanks to ordinary least square multivariate modelling and classifying countries is using principal component analysis, K-means, and hierarchical clustering. The study demonstrates that there are correlations between growth parameters directly linked to the occurrence of new cases of COVID-19 and socio-economic variables, specifically current health expenditure as a percentage of gross domestic product (CHE/GDP) is anti-correlated with basic reproduction number, demonstrating the effectiveness of public health mitigation measures, even if they involve significant medico-economic costs. The result reveals a disparity between developed and developing countries, despite the fact that rapid implementation of isolation and vaccination measures allowed them to anticipate and mitigate the effects of the second wave. Section 4.5 deals with the techniques of functional data analysis applied to model daily hospitalized, deceased, Intensive Care Unit (ICU) cases, and return home patient numbers along the COVID-19 outbreak, considered as functional data across different departments in France, while response variables are numbers of vaccinations, deaths, infected, recovered, and tests numbers. The first functional principal component well characterized the forms seen for the data connected to the first three principal components, according to the functional principal component analysis. This discovery confirms the significance of the first component in explaining and qualitatively predicting the observed data. The influence of vaccination is visible because the shapes are attenuated after vaccination and do not resemble the shapes observed in Section 3.2 for seasonal influenza. Finally, this chapter is concluded in Section 4.6 with some perspectives and conclusions.

4.1 Data Analysis and Forecasting of COVID-19 Pandemic Based on Daily Observation Dynamics

This section deals with the application of several statistics tools for modelling epidemic data such as ARIMA model, Exponential smoothing model, Holt's method, Prophet forecasting model and machine learning models. Among machine learning models, log-linear, polynomial and support vector regressions were used in order to compare these models. The main objective and motivation is to present the best statistical model to describe the daily count of new cases and deaths due to COVID-19 outbreak. The best model is therefore used to predict the number of new cases and to model the mortality due to the COVID-19 pandemic. Another objective is to decompose the COVID-19 data into two main waves and model each wave separately using different tools along with the classical time series model that depends on the basic reproduction number R_o, followed by many different additive models to predict the time series such as Exponential regression, log-linear and Polynomial with different degrees. In addition, several smoothing methods, Prophet and Support vector methods were applied

to COVID-19 data. A comparison is made between these different techniques for the best fit. These different tools have been applied to some countries data and the models are supported by validity and accuracy tests. The sensitivity analysis for the parameters used in the ARIMA model and exponential models are presented.

4.1.1 ARIMA modeling for first wave and second wave

This section deals with the application of ARIMA model to calculate the initial autocorrelation slope and also to forecast new cases for both waves for some countries. The ARIMA model shows more than 95% confidence interval as it can be seen in Figure 4.1(a)-(d) with p-value for Mali for first wave is $p = 0.01$ and for second wave it is $p = 6.3 \times 10^{-10}$ while for first wave for Slovenia $p = 0.01$ and for second wave in Luxembourg $p = 0.01$.

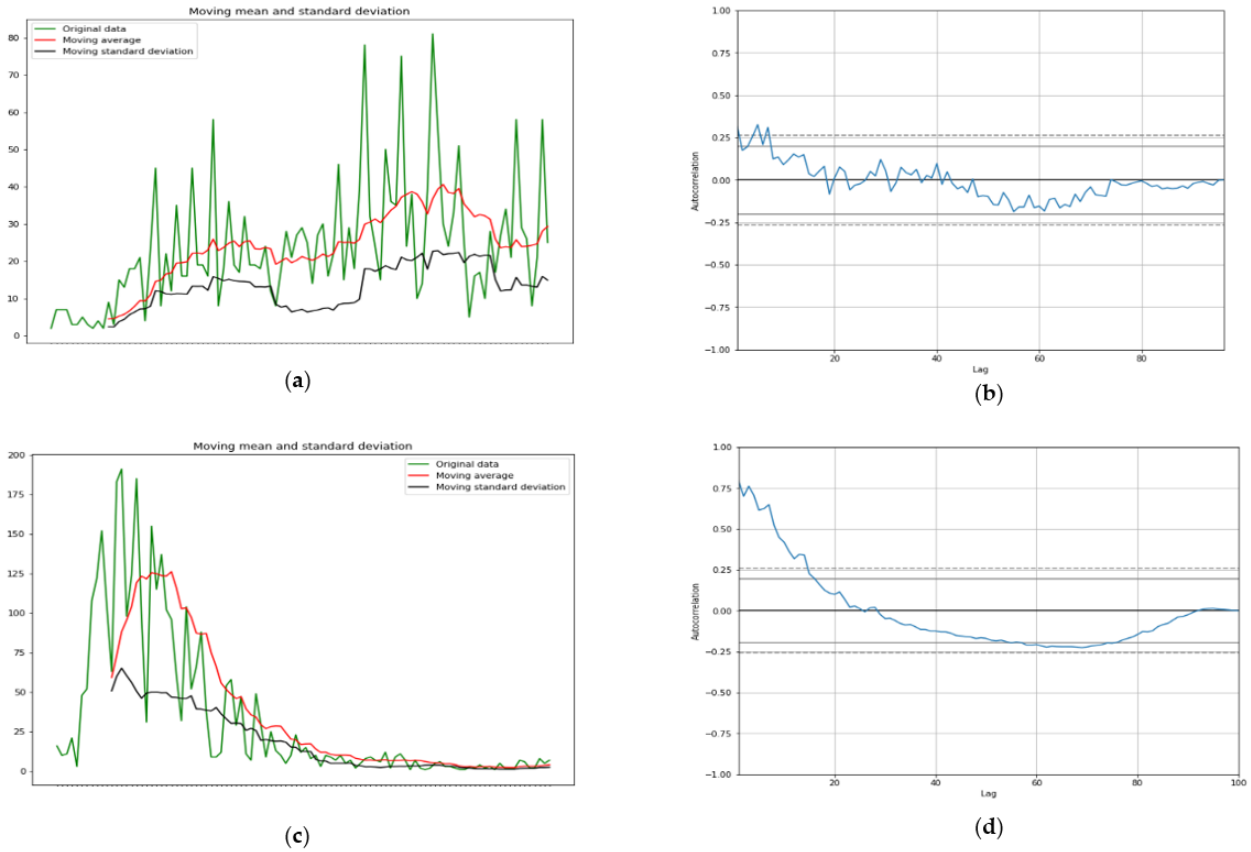


Figure 4.1: (**a**) First wave moving average and standard deviation of new cases (left) plotted against time in days on x-axis and (**b**) autocorrelation curve for Mali (right). (**c**) First wave moving average and standard deviation of new cases (left) plotted against time in days on x-axis and (**d**) autocorrelation curve for Luxembourg (right).

Figure 4.1 compares two countries, one from Sahelian Africa, Mali and one from western Europe, Luxembourg during the first wave of COVID-19 outbreak during the spring 2020: Mali shows a quasi-endemic behavior with a weakly varying autocorrelation function and Luxembourg a frank epidemic wave with a classic shape. For the second wave in fall 2020, Mali presents an attenuated epidemic shape (due probably to specific geoclimatic conditions in western Africa [102]) and a country from central Europe, Slovenia, shows at this period

an endemic behavior with an oscillatory occurrence of new cases. Figure 4.3 proposes a forecasting based on ARIMA decomposition for the first and second waves in Mali with a better approximation for the epidemic second wave than for the quasi-endemic first wave. It is the same for Luxembourg with an inversion of the phases order, an epidemic wave followed by an endemic state well predicted. On the contrary, for Slovenia, the endemic state with oscillations is badly predicted.

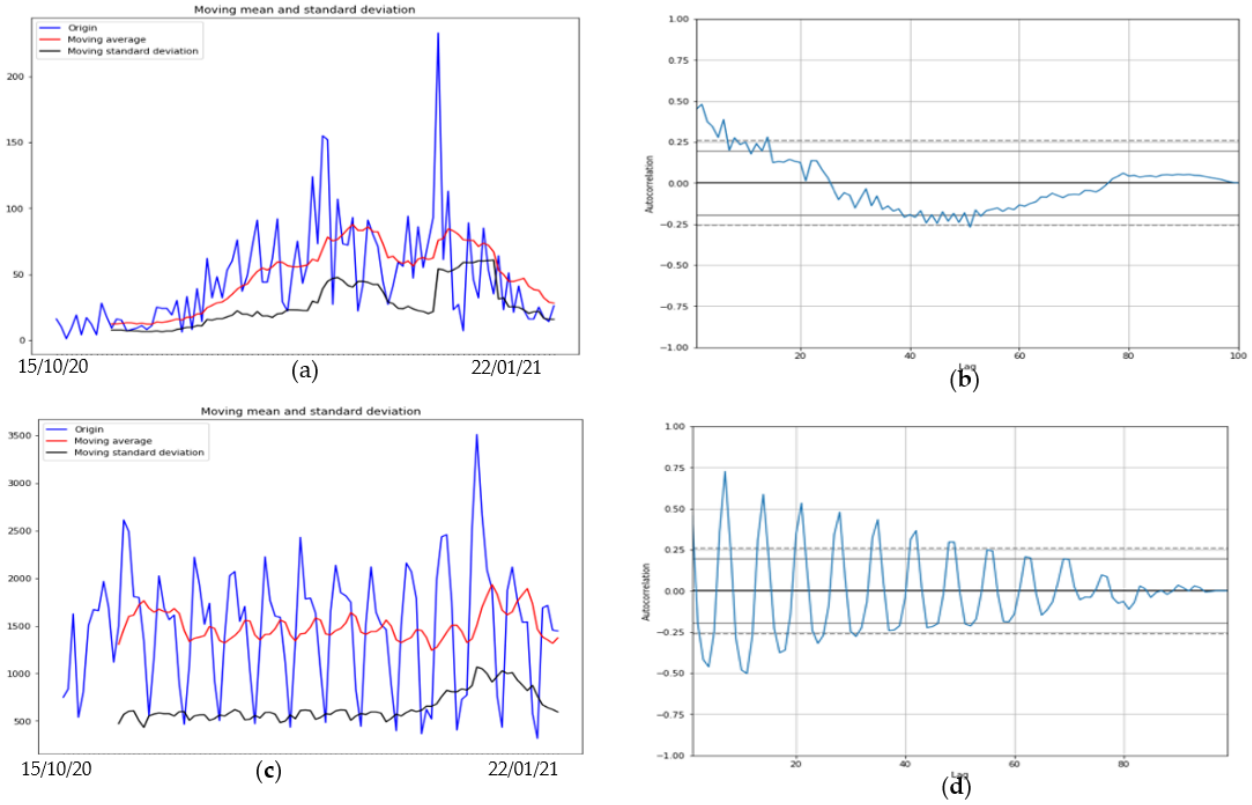


Figure 4.2: (**a**) Second wave moving average and standard deviation of new cases (left) plotted against time in days on x-axis and (**b**) autocorrelation curve for Mali (right). (**c**) Second wave moving average and standard deviation of new cases (left) plotted against time in days on x-axis and (**d**) autocorrelation curve for Slovenia (right).

The comparison during the first wave between two countries (Figure 4.1), one developed (Luxembourg) and one developing (Mali) shows a difference in length of contagiousness period (linked to the value of the opposite to the slope at origin of the autocorrelation function) and shape of the growth curve, indicating a lower virulence of the SARS Cov-2 in Mali, possibly due to the influence of the temperature [102]. This tendency is reversed during the second wave between Mali and Slovenia (Figure 4.2). The forecast using the ARIMA method shows a good retrospective adjustment to past data, but a weak predictive power of the future trend of new cases, in particular for the prediction of the entry into the endemic phase after an epidemic wave (Figure 4.3).
The initial negative autocorrelation slope of the epidemic spread averaged on six days is also calculated and presented for Kuwait while other countries is summarized in Appendix B, Table B.1. As expected, the 100 first days of the first wave start show a positive slope (0.0687) and the 100 first days of the second wave transition show a negative slope (-0.0094). The results is compared using parameters (p,d,q) and the best result gave lower RMSE and

p-value. We trained 60% of the data and forecast 100 days while the rest was tested to show the validation and accuracy of the model. It was observed that while the RMSE of the tested data was high, by checking the residual using Ljung-Box test, the Q^* is large with p-value < 0.05 which shows that the autocorrelation did not come from white noise. The visualisation result is in Figure 4.4: for first wave new cases the ARIMA parameters used are (0,1,1) with drift, RMSE for training = 30.7, RMSE for test = 533.5, $Q^* = 33$ and p-value= 0.00007. For second wave transition new cases the ARIMA parameters used are (0,1,2) with drift, RMSE for training = 74.5, RMSE for test = 402.3, $Q^* = 15$ and p-value= 0.04.

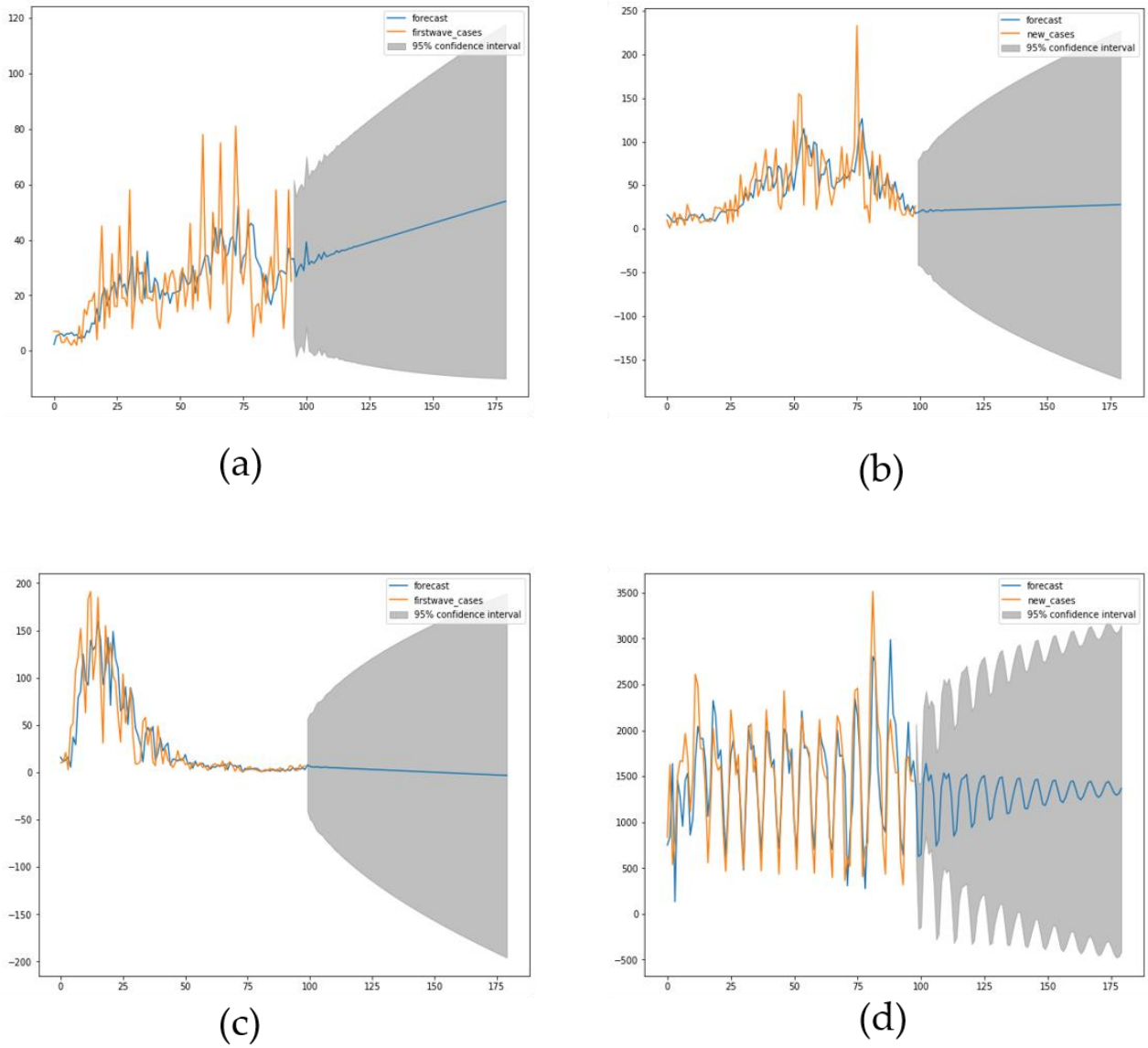


Figure 4.3: (**a**) First and (**b**) second wave forecast for Mali. (**c**) First wave forecast for Luxembourg. (**d**) Second wave forecast for Slovenia. All plotted against time in days on x-axis.

Using the Death data we apply the ARIMA model and got: for first wave deaths the ARIMA parameters used are (1,1,1), RMSE for training = 2.2, RMSE for test = 2.37, $Q^* = 16$ and p-value= 0.04. For the second wave transition deaths the ARIMA parameters used are (0,1,0) and this is the only case where the p-value was high despite low RMSE:

the model degree of freedom is zero, RMSE for training = 2.21, Root MSE for test = 0.894, $Q^* = 13$ and p-value= 0.2. It can be concluded that the ARIMA model best fits the first wave deaths with the best performance.

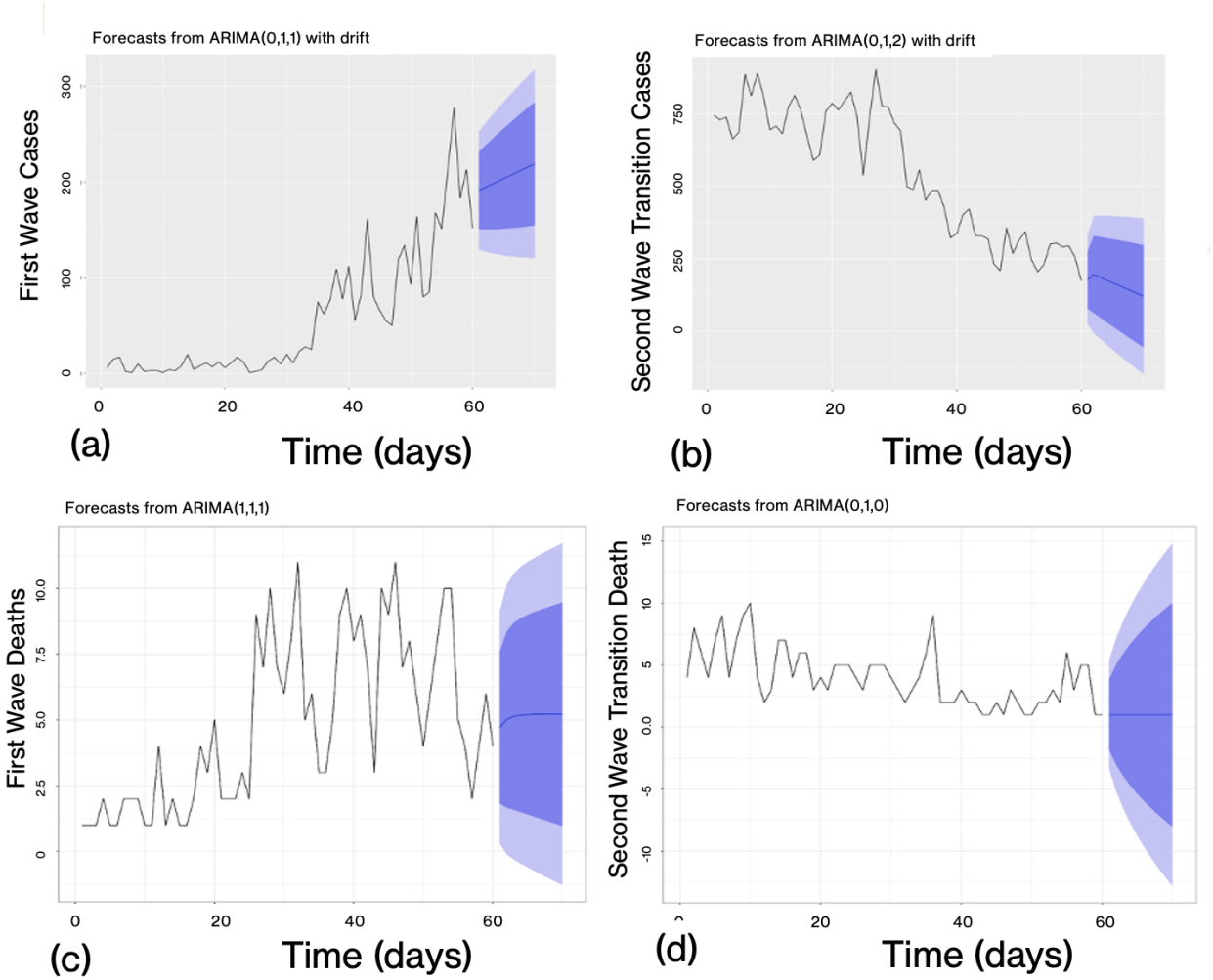


Figure 4.4: (**a**) ARIMA forecast for first wave new cases in Kuwait. (**b**) ARIMA forecast for second wave transition cases in Kuwait. (**c**) ARIMA forecast for first wave deaths in Kuwait and (**d**) ARIMA forecast for second wave transition deaths in Kuwait.

4.1.2 Regression analysis for first wave and second wave

In this Section, regression models are applied on the daily new cases for both waves observed in Kuwait and it is also applied to some countries epidemiologic variables (maximum reproduction number, exponential slope calculated in this section and the initial autocorrelation slope calculated in the previous section) in order to establish the relationship among them. Looking at the pattern of COVID-19 data of Kuwait, one realises that it consists of two main waves. Each wave lasts approximately three months. Based on available data, 100 days both in the first wave and second wave was considered. First wave new cases were considered from 25/02/2020 to 03/06/2020 while second wave new cases were considered from 15/10/2020 to 22/01/2021. For the first wave, daily deaths data is from 04/04/2020 to 12/07/2020, while for the second wave is from 15/10/2020 to 22/01/2021. The slopes from the log-linear regression

analysis is calculated using exponential model. The result of exponential slope for Kuwait is presented in this section while other countries is summarized in Appendix B, Table B.1.

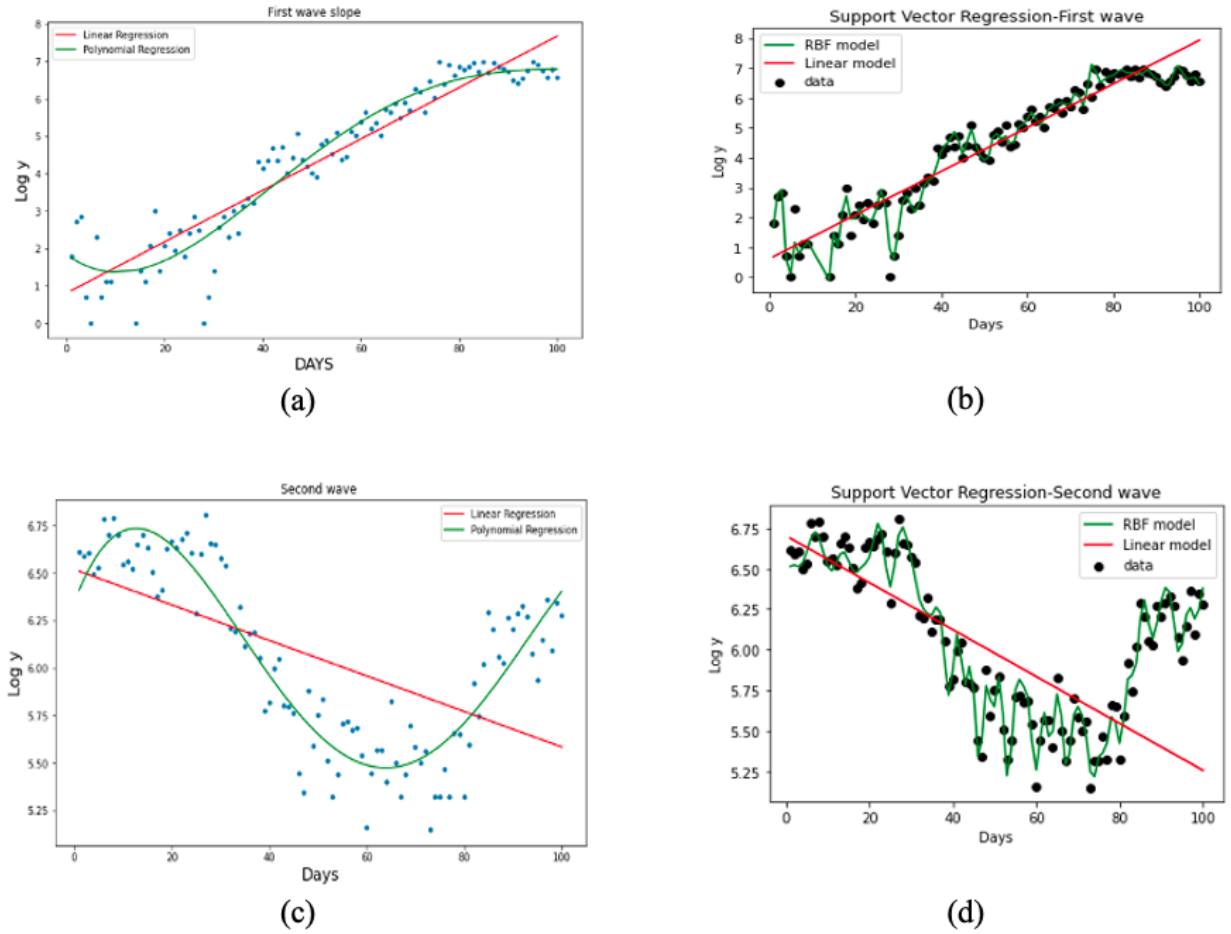


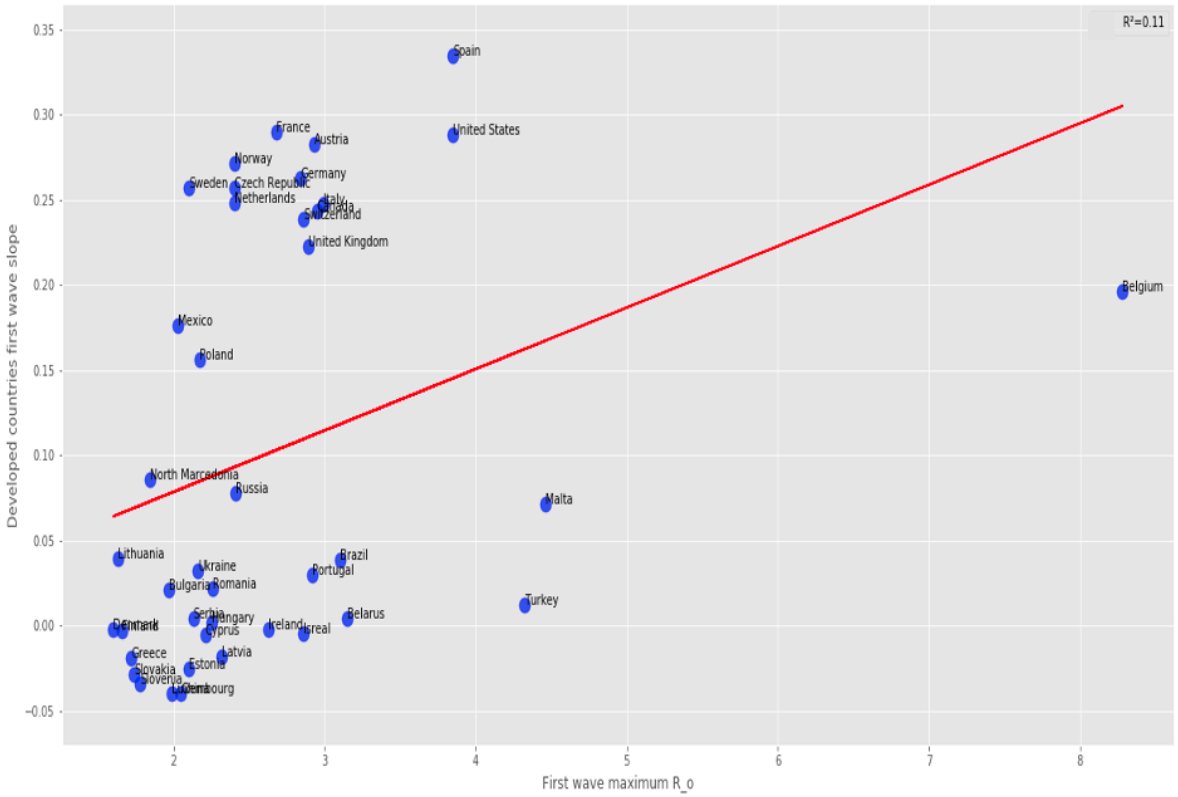
Figure 4.5: (**a**) Log-linear and polynomial regression for first wave cases in Kuwait. (**b**) Support vector regression for first wave cases in Kuwait. (**c**) Log-linear and polynomial regression for second wave transition cases in Kuwait and (**d**) Support vector regression for second wave transition cases in Kuwait. The origin of time corresponds to the corresponding wave start.

For first wave, the exponential model: $y = 0.8e^{0.07x}$, based on the following results:

- Log-linear regression: slope = 0.0686852051631, intercept=0.800565672854, $r = 0.939152595486$, p-value = $2.079801359e^{-45}$, standard error = 0.002591135543915, R-squared = 0.882008, RMSE = 0.711400960815.
- Using polynomial regression of order 4 we have: R-squared = 0.787421, RMSE = 0.586403581239, p-value< $2.2e^{-16}$, standard error = 0.6069, F-statistics = 256.7.
- Using Support vector regression we get: R-squared = 0.98382, RMSE = 0.263424533991, slope = 0.0128466, intercept= 3.697591, p-value< $4.33e^{-15}$.

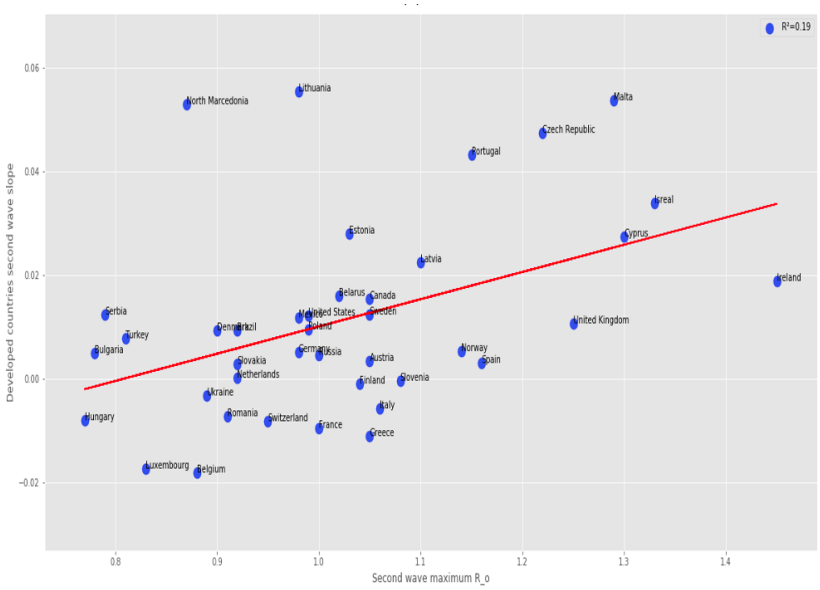
For second wave transition, the exponential model is $y = 6.5e^{-0.009x}$, based on the following results:

- Log-linear Regression: slope = -0.0093674219195, intercept = 6.5198283885, r = -0.57106504118, p-value = 5.50386624132e-10, standard error = 0.00136023603, R-squared = 0.326115, RMSE = 0.388700386316,
- Polynomial of order four Regression: R-squared = 0.0927367, RMSE = 0.1818719990, p-value$< 2.2e^{-16}$, standard error = 0.186, F-statistics = 137.2,
- Support Vector Regression: R-squared = 0.9680774158, RMSE = 0.0846001608, slope = -0.2570435, intercept= -1.370541, p-value = 0.2775.



(a)

Figure 4.5 show different regressions for the first and transition to second waves of COVID-19 in Kuwait. All coefficients for the first wave and transition to second wave for both Log-linear and polynomial regression are significant with p-value less than 0.001. The residuals of the log-linear regression were examined and it was discovered that the median for both the cases were close to zero with median = 0.06031 for the first wave and median = 0.07629 for transition to the second wave. The normality of the residual was tested using Jarque-Bera test and with high p-value we fail to reject the null hypothesis that the skewness and kurtosis of the residuals are significantly equal to zero. Also, the median of the residual for both first wave and transition to second wave of polynomial regression of order four is close to zero with median = 0.0271 and median = 0.00669 respectively. For support vector regression, we tested the normality of the residual using the Jarque-Bera test and it was discovered that it's skewness and kurtosis are significantly equal to zero since the p-value is large.
We now apply linear regression model by using some historic data (100 days infectivity period in this case) on some countries epidemiologic variables and consider a linear relationship



(b)

Figure 4.6: Regression plots for developed countries of (**a**) first and (**b**) second wave maximum R_0 of the new cases curve.

between them while polynomial regression models use a similar approach but the dependent variable is modeled as a degree n ($6 \geq n \geq 2$) polynomial in x. It is presented as follows:

- Linear Regression
  - Figure 4.6(a): slope = 0.03612, intercept = 0.0062, r-value = 0.3299, p-value = 0.0352, standard error = 0.0165, R-squared = 0.109, RMSE = 0.116.
  - Figure 4.6(b): slope = 0.05223, intercept = -0.0421, r-value = 0.434366, p-value = 0.0051, standard error = 0.01757, R-squared = 0.18867, RMSE = 0.01689.
  - Figure 4.6(a) and Figure 4.6(b) show a positive correlation between the slope of the logarithmic regression curve of the new cases of COVID-19 as a function of time (a sign of rapid growth of the epidemic if it is high). The early implementation of mitigation measures in developed countries helps in reducing the exponential growth of new cases in the second wave. This trend is confirmed in the study of the correlation between the slope of the logarithmic regression and the maximum R_0 (Figure 4.6(a) and Figure 4.6(b)), which increases during the second wave in developed countries (the correlation coefficient rising from 0.33 to 0.44), showing a growth of the new cases more brutal, but shorter, undoubtedly due to the establishment of a faster and more effective lock-down.

- Parabolic and Cubic Regression
  - Figure 4.7 show the classical linear and polynomial regressions (parabolic for Graphs (a) and (b) and cubic for Graph (c)) between the opposite of the slope at the origin of the autocorrelation function of the ARIMA model and successively the slope of the logarithmic regression line of the new daily cases of COVID-19 of the first wave (a), then that of the second wave (b), and finally the number of days since the start of the outbreak (c).

More precisely, on Figure 4.7(a)-(b) first and second waves of the COVID-19 pandemic are compared using linear and parabolic or cubic regression, showing a significant positive (resp. negative) correlation between the opposite of the initial autocorrelation slope and exponential regression slope of the first (resp. second) wave for developed (resp. all) countries. This opposition between the two waves could result from the application of a more severe lockdown in developed countries during the second wave. On Figure 4.7(c), the opposite of the initial autocorrelation slope decreases significantly if the start of the first wave in a country is late with respect to the start of the COVID-19 outbreak in China due probably to the progressive implementation of mitigation measures in that country taking into account the experience of the countries starting first wave before.

- For Figure 4.7(**a**):

- Linear regression: slope $= -0.193$, intercept $= 0.102$, r-value $= -0.394$, p-value $= 1.026 \times 10^{-7}$, standard error $= 0.03467$, RMSE $= 0.0385$,
- Polynomial regression (order 2): p-value $= 0.00145$, standard error $= 0.54339$, R-squared $= 0.19$, , RMSE $= 0.046$.

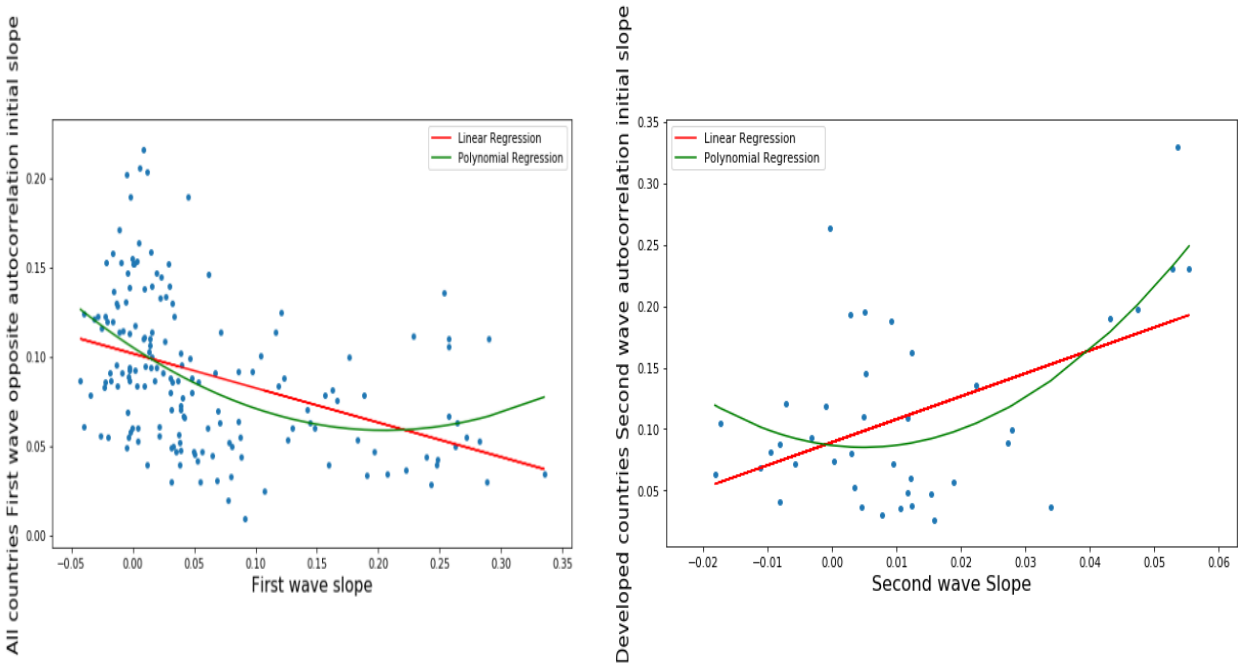
- For Figure 4.7(**b**):

- Linear regression: slope $= 1.867$, intercept $= 0.089$, r-value $= 0.487$, RMSE $= 0.063$,
- Polynomial regression (order 2): R-squared $= 0.37$, RMSE $= 0.094$.

- For Figure 4.7(**c**):

- Linear regression: slope $= 0.000295$, intercept $= 0.0765$, r-value $= 0.195469$, p-value $= 0.01415$, standard error $= 0.000119$, R-squared $= 0.038$, RMSE $= 0.04$,
- Polynomial regression (order 3): R-squared $= 0.1$, RMSE $= 0.0414825$.

The curves show a different behavior between the two waves (a) and (b), probably due to an increase in the contagion parameter, the basic reproduction number R_0 (linked to the Malthusian parameter of the exponential growth phase), despite a shortening of the duration of contagiousness (linked to the slope at the origin of the auto-correlation function, which is all the stronger as the distance from the start of the epidemic increases, no doubt because of the mitigation measures, which decrease the duration of the contagiousness period).



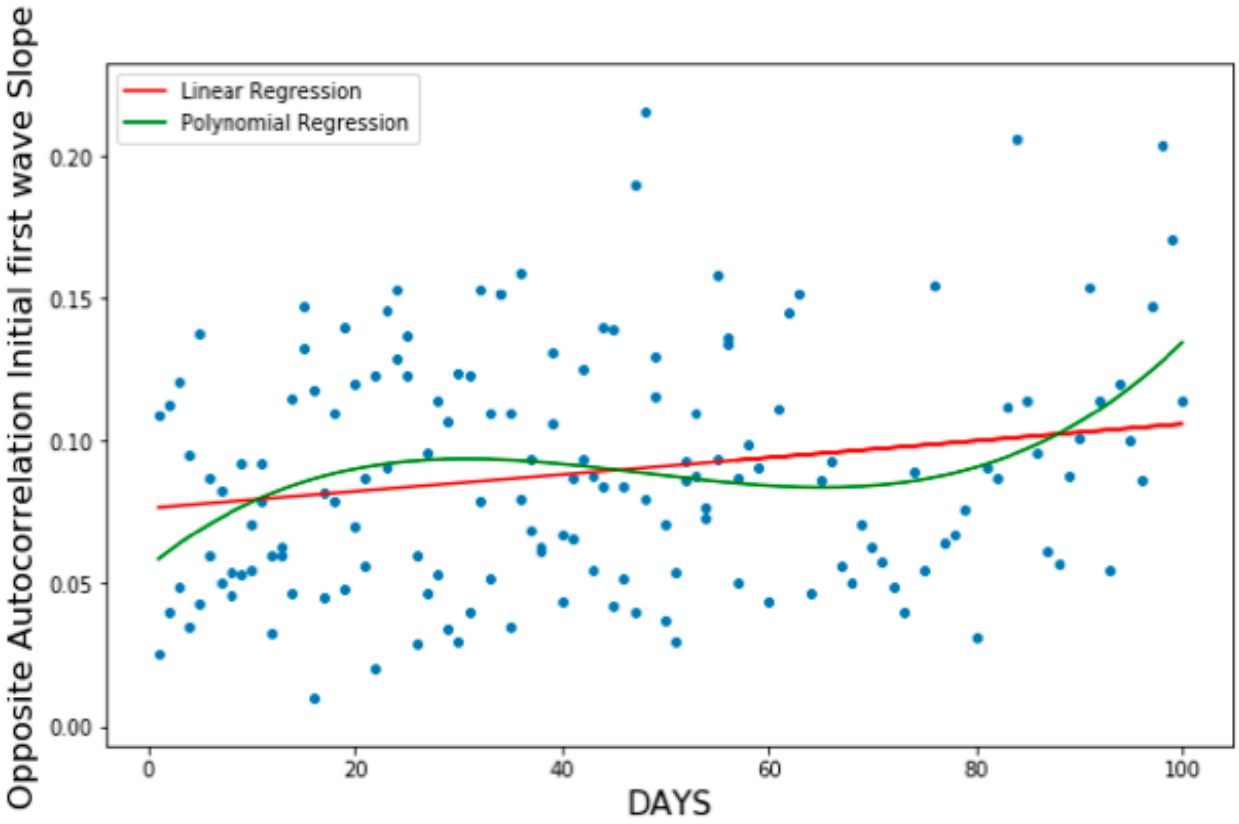
(**a**) (**b**)

- Quartic Regression

  Figure 4.8 is obtained by using a polynomial of degree 4, a fit showing a minimum for the value of the maximum R_0 equal to 3.5, which is considered as the observed value for the maximal effective reproduction number at start of the first wave in many developed countries (France, Germany, Switzerland, UK, USA, etc.) [126], which corresponds to the fact that the opposite of the initial autocorrelation slope (indicating that the length of the contagiousness is short when the absolute value of the slope is high) decreases (the contagiousness duration increases) when the maximum R_0 increases, which seems logical. The result is as follows:

  - Linear regression: slope = 0.01034, intercept = 0.1019, r-value = -0.3578, p-value = 0.02163, standard error = 0.00433, RMSE = 0.0303,
  - Polynomial regression (order 4): RMSE = 0.0349, R-squared = 0.33.

    On Figure 4.8, the opposite of the initial autocorrelation slope is significantly negatively correlated with the maximum R_0 observed at the inflection point of the new cases curve, confirming that long contagiousness periods give high exponential increases of the new cases.



(c)

Figure 4.7: Linear (in red) and parabolic or cubic (in green) regression plots of the opposite of the initial autocorrelation slope vs. (**a**) first wave exponential regression slope for all countries, (**b**) second wave exponential regression slope for developed countries and (**c**) days from the start of the first wave observed for all countries.

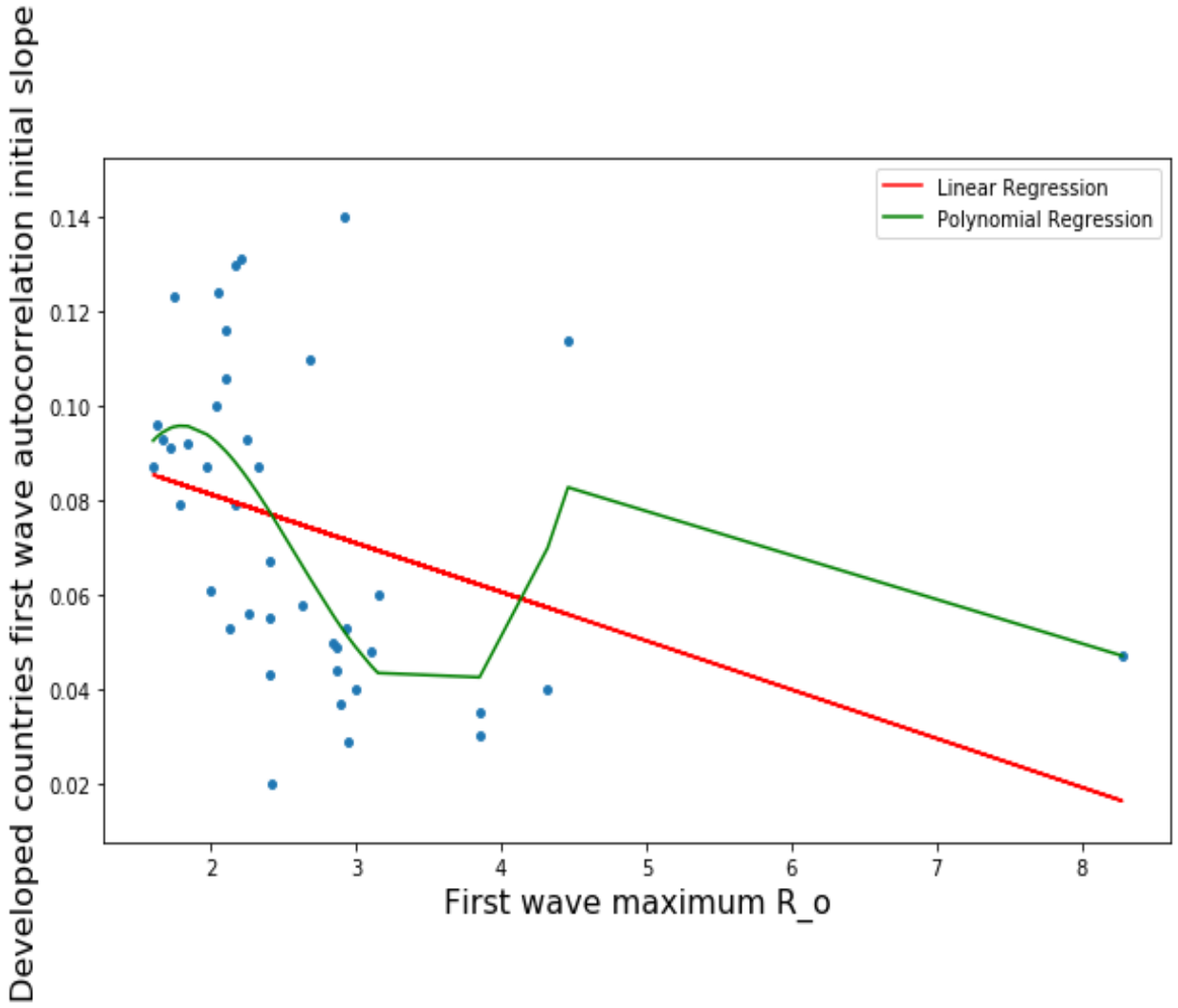


Figure 4.8: Linear (in red) and quartic (in green) regression plots of the opposite of the initial autocorrelation slope of the first wave vs first wave maximum R_0 for developed countries.

4.1.3 Performance, accuracy and validation of the regression models

In order to know the performance of the regression models we trained 80% of the data from Kuwait and test 20% percent of the data and also did cross validation to be sure of the accuracy. The predicted and the observed values are very close to the result presented for all the regression models. For Log linear model we present the cross validation result in Figure 4.9(a) whose average mean square errors for the 5 portion folds are 0.5027791 for first wave on the left and 0.1533665 for second wave transition on the right. We observed high correlations between the tested and the predicted values (R-squared = 0.9278587 for first wave cases and R-squared = 0.5312499 for second wave transition cases) for both cases.
For polynomial regression of order four, we present the performance of the test model as follows: for first wave cases, multiple R-squared= 0.9437, p-value= $3.448e^{-09}$, relative standard error is 0.5152 and the residual median is 0.03408, value close to zero which shows that the model performs very well; for second wave transition cases, multiple R-squared= 0.9039, p-value= $1.829e^{-07}$, relative standard error is 0.194 and the residual median is 0.00386, value close to zero which shows that the model performs optimally.

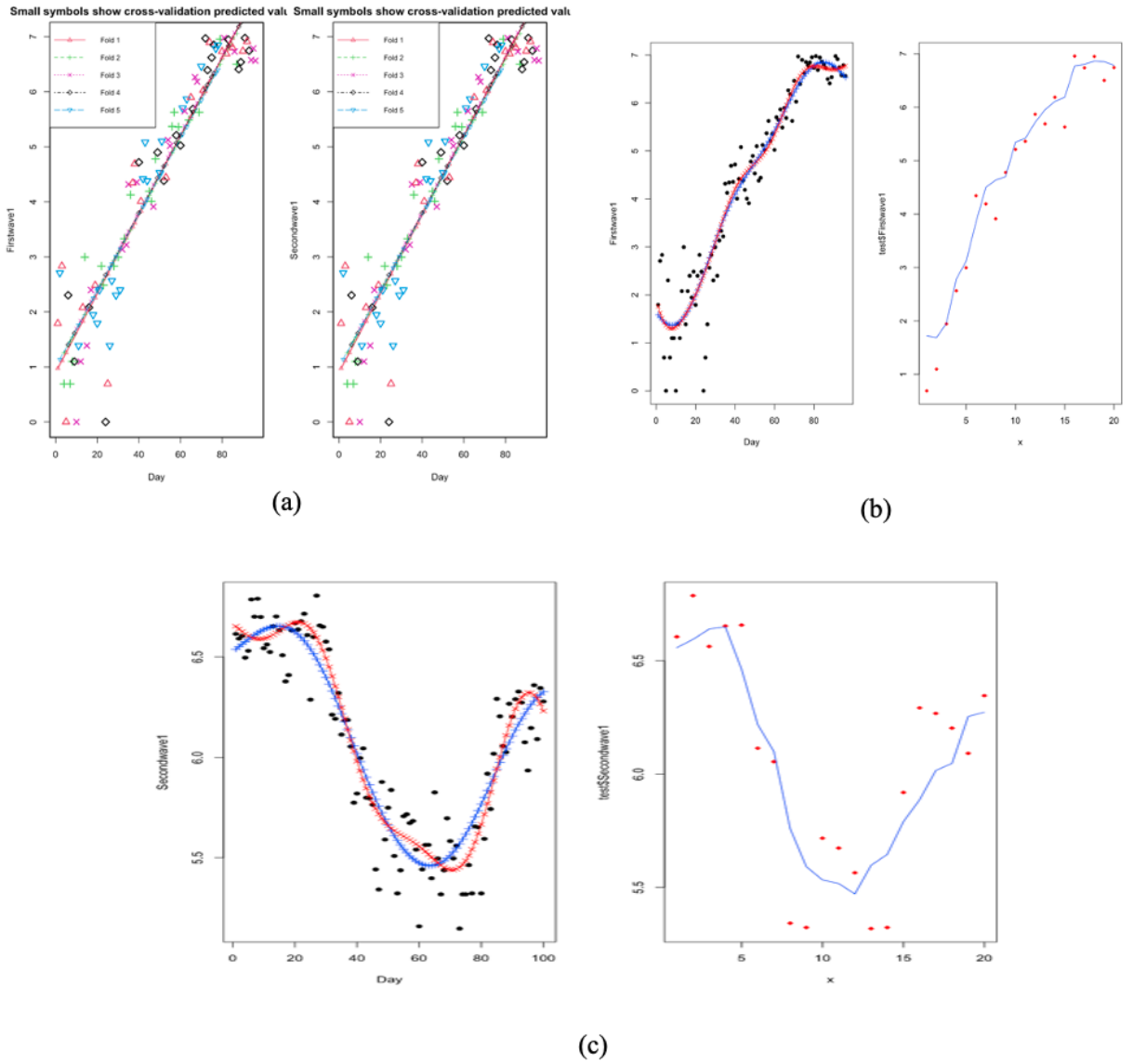


Figure 4.9: (a) Cross validation result for Log linear model for new cases in Kuwait. (b) Comparison of the tuned support vector model and support vector regression model with the test prediction for first wave new cases in Kuwait. (c) Comparison of the tuned support vector model and support vector regression model with the test prediction for second wave transition new cases in Kuwait.

Lastly for support vector regression we present the optimum model with parameters ($\epsilon = 0$ for both cases, and cost $= 4$ and cost $= 10$) respectively, for the first and second wave transition cases in Figure 4.10(a) and Figure 4.10(b) with mean square error values of 0.3840138 and 0.03 respectively using 10 folds cross validation. RMSE for the first and second wave transition cases are 0.5757138 and 0.16 respectively, slope values of -0.0452387 and -0.1503307 respectively, p-value $= 1.554e^{-15}$ for first wave cases, p-value $= 0.293$ for second wave transition cases and intercept values 3.751997 and 3.85671 respectively. We present a comparison of RMSE for support vector model, tuned support vector model, constructed support vector model and also the test model in Figure 4.9(b) for first wave cases and Figure 4.9(c) for second wave transition cases. The left hand side of both figures is the tuned support vector

model while the right hand side is the test model prediction. The test model performance is presented as follow: first wave cases: R-squared = 0.8450823 and RMSE = 0.784234; second wave transition cases: R-squared= 0.8516843 and RMSE = 0.1668129.

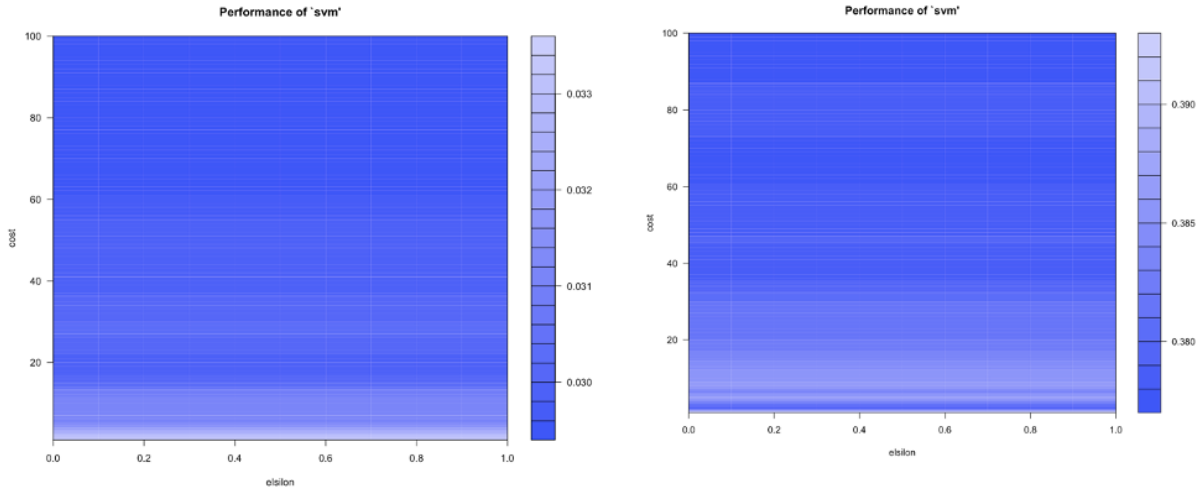


Figure 4.10: (a) Optimum support vector model mean square error visualisation for first wave new cases in Kuwait and (b) Optimum support vector model mean square error visualisation for second wave transition new cases in Kuwait.

4.1.4 Exponential smoothing method

Simple exponential model (SES or ANN) is a good forecasting tool and in this case we used $\alpha = 0.2$ as one of the parameters while other parameters varies depending on the case in consideration. The degree of freedom (df) is 8 while the model degree of freedom is 2 with 10 lags. Also, 60% of the Kuwait data was trained with 100 days forecasting while 40% was tested. It was observed that while training and testing the model, the best exponential model is when the trend is removed from the test set as it can be seen in Figure 4.11 even though the test RMSE is quite large but the p-value is low. Other exponential models like diff.SES (removing trend from SES), simple exponential smoothing with multiplicative errors (MNN) and simple exponential smoothing with additive errors (MAN) did not give better performance when their RMSE was compared which were extremely larger than the simple exponential model without trend. The median of the residual is zero for all cases except first wave cases whose median is -55, which shows that residual is normally distributed. The parameters and result for each case is as follows: for first wave cases the parameters used are $l = 0.48$ and $\sigma = 39$, RMSE for training = 38.3, RMSE for test = 219.9, $Q^* = 78$ and p-value= $1e^{-13}$; for second wave transition cases the parameters used are $l = 0$ and $\sigma = 92.3$, RMSE for training = 90.7, RMSE for test = 76.2, $Q^* = 30$ and p-value= $2e^{-04}$; for first wave deaths the parameters used are $l = 0.0106$ and $\sigma = 2.73$, RMSE for training = 2.69, RMSE for test = 2.07, $Q^* = 29$ and p-value= $4e^{-04}$ and for second wave transition deaths the parameters used are $l = 0.4891$ and $\sigma = 2.49$, Root MSE for training = 2.45, RMSE for test = 1.05, $Q^* = 14$ and p-value= 0.09. It can be concluded that the simple exponential smoothing model without trend works better for first wave deaths.

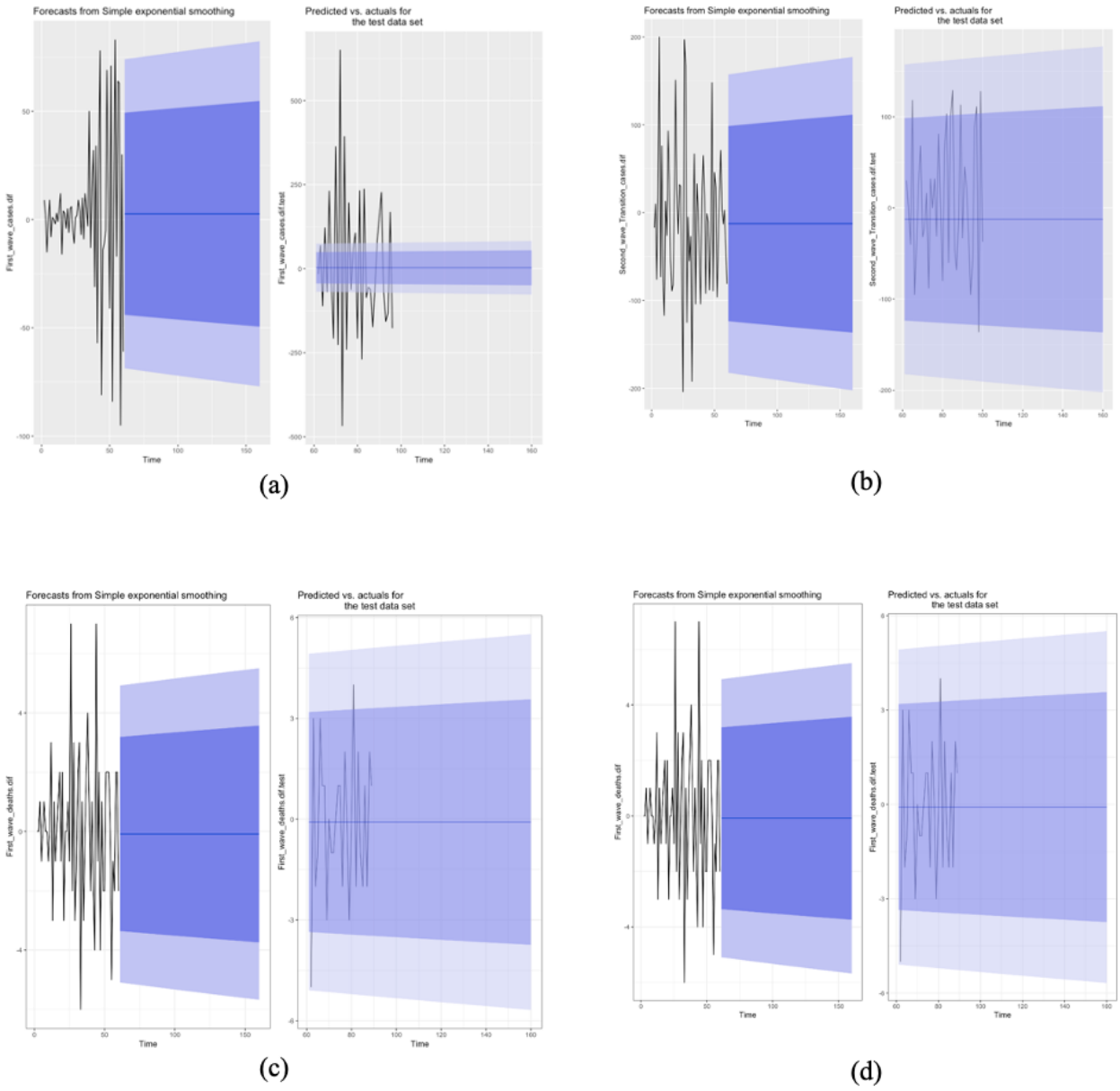


Figure 4.11: (a) Simple exponential smoothing for first wave cases in Kuwait. (b) Simple exponential smoothing for second wave transition cases in Kuwait. (c) Simple exponential smoothing for first wave deaths in Kuwait and (d) Simple exponential smoothing for second wave transition deaths in Kuwait.

It is also important to present the visualisation of different types of exponential smoothing models (see Figure 4.12(a)-(d)) to give significance for recent observations and produce accurate forecasts of 100 days while Holt's model (AAN) as shown in Figure 4.13(a)-(d) gives the trend and level of a time series and is computationally more efficient than double moving average. Holt's-Winters' model considers randomness using efficient smoothing process and is computationally efficient too. Holt's linear method with additive errors did not give better performance when we compared with the optimal Holt's model because the RMSE was extremely larger.

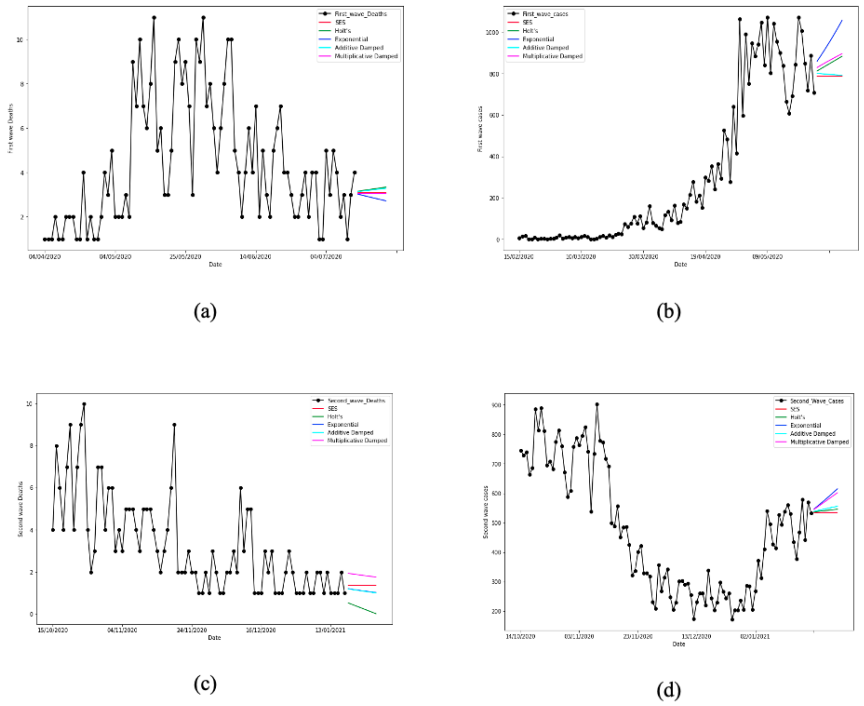


Figure 4.12: Comparison between different exponential smoothing methods for (a) first wave deaths, (b) first wave cases, (c)second wave transition deaths and (d) second wave transition cases in Kuwait.

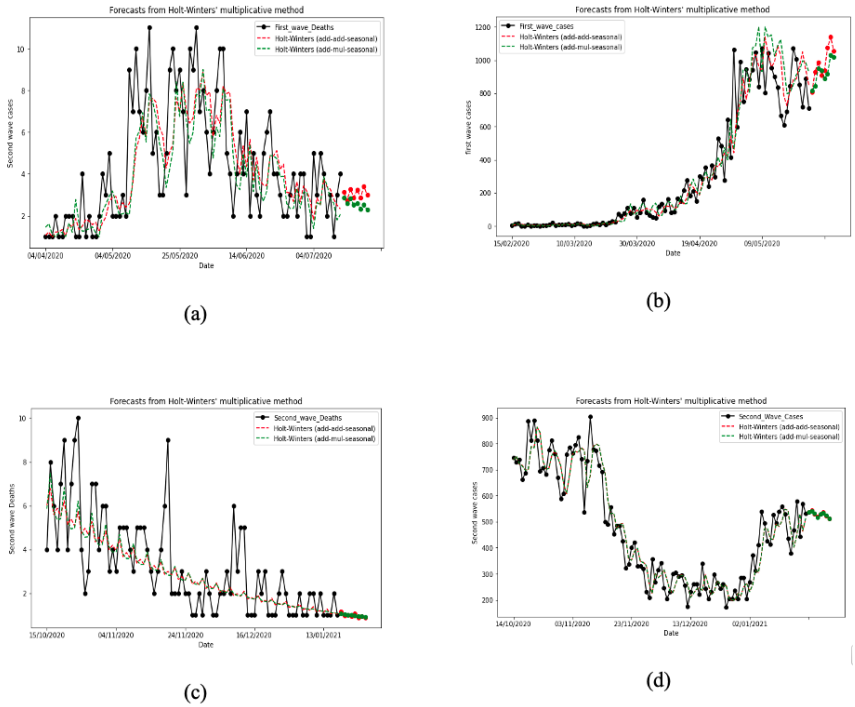


Figure 4.13: Holt method for (a) first wave deaths, (b) first wave cases, (c)second wave transition deaths and (d) second wave transition cases in Kuwait.

4.1.5 Performance, accuracy and validation of the exponential models

We give an explicit analysis of the performance and accuracy of the simple exponential without trend and Holt's model using Kuwait data. Figure 4.15(a)-(d) gives the comparison between optimal RMSE's of simple exponential model without trend and Holt's model. We trained 60% of the model while others were tested and 10 lags were used for the modelling. Figure 4.14(a)-(d) show visualisation of comparison between the holts model and the optimal Holt's model

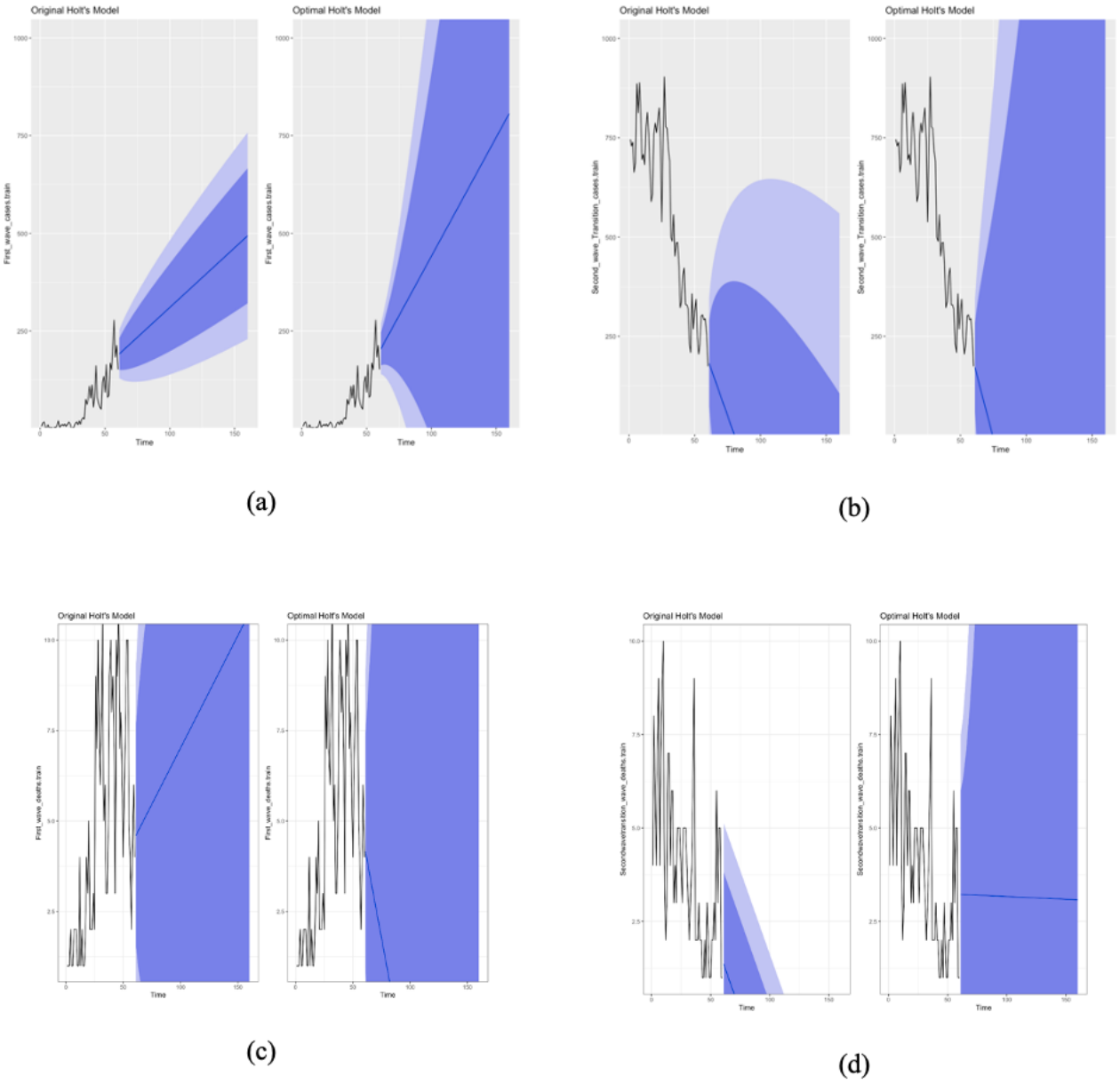


Figure 4.14: Comparison of Holt's model and optimal Holt's model forecast for (a) first wave cases, (b) second wave transition cases, (c) first wave deaths and (d) second wave transition deaths in Kuwait.

- for first wave cases : Optimal model for simple exponential model without trend parameters are $\alpha = 0.05, l = 1.3$ and $\sigma = 36.4$, with result of RMSE for training =

35.8, RMSE for test = 219.4, Holt's model parameters are $\alpha = 0.04095, l = 8.9848, b = 3.0532, \beta = 0.0004$, df= 6, model df = 4 and $\sigma = 31.8$, with result of RMSE for training = 30.7, RMSE for test = 534.4, $Q^* = 32$ and p-value= $1e^{-05}$, Holt's optimal model result of RMSE for training = 21.3, RMSE for test = 409.3;

- for second wave transition cases : optimal model for simple exponential model without trend parameters are $\alpha = 0.05, l = -5.0588$, df= 8, model df = 2 and $\sigma = 86.7$, with result of RMSE for training = 85.2, RMSE for test = 73.2, $Q^* = 30$ and p-value= $2e^{-04}$, Holt's model parameters are $\alpha = 0.7798, l = 755.866, b = -9.4126, \beta = 0.0004$, df= 6, model df = 4 and $\sigma = 84.9$, with result of RMSE for training = 82, RMSE for test = 422, $Q^* = 35$ and p-value= $4e^{-06}$, Holt's optimal model result of RMSE for training = 84.6, RMSE for test = 504.5;

- for first wave deaths : Optimal model for simple exponential model without trend parameters are $\alpha = 0.05, l = 0.107$, df= 8, model df = 2 and $\sigma = 2.55$, with result of RMSE for training = 2.51, RMSE for test = 2.06, $Q^* = 27$ and p-value= $6e^{-04}$, Holt's model parameters are $\alpha = 0.5016, l = 0.6829, b = 0.0621, \beta = 0.0004$, df= 6, model df = 4 and $\sigma = 2.38$, with result of RMSE for training = 2.3, RMSE for test = 271, $Q^* = 27$ and p-value= $1e^{-04}$, Holt's optimal model result of RMSE for training = 2.40, RMSE for test = 2.46;

- for second wave transition deaths : Optimal model for simple exponential model without trend parameters are $\alpha = 0.05, l = 0.0422$, df= 8, model df = 2 and $\sigma = 2.32$, with result of RMSE for training = 2.28, RMSE for test = 0.94, $Q^* = 13$ and p-value= 0.1, Holt's model parameters are $\alpha = 0.0004, l = 6.8362, b = -0.0891, \beta = 0.0004$, df= 6, model df = 4 and $\sigma = 1.92$, with result of RMSE for training = 1.85, RMSE for test = 1.50, $Q^* = 8$ and p-value= 0.2, Holt's optimal model result of RMSE for training = 1.79, RMSE for test = 1.79.

To conclude this Section, it is very important to give a critical look at the RMSE comparison in Figure 4.15a-d because it guides the choice of parameters to be used in the model. In Figure 4.15a we observed that for simple exponential model without trend on the left, if the α value is increased beyond 0.2, the RMSE value is on the increasing trend while for Holt's model on the right, it is best to use β of about 0.05 and after then the RMSE is on the increasing trend. In Figure 4.15b we notice that for the simple exponential model without trend on the left the choice for α must be below 0.2 and after that the RMSE continues to increase while for Holt's model on the right, β must be below 0.1 else the RMSE value becomes large. The Figure 4.15c is a bit tricky because for the choice of α makes the values of RMSE to be stationary till the point where α is 0.6 where the increasing trend begins for simple exponential model without trend on the left while for Holt's model on the right there was turning point when the value of β is 0.15 but the value gives minimum RMSE when $\beta = 0.04$ and $\beta = 0.38$, which means that the choice of β for this case is very critical and must be precise. In Figure 4.15d, the simple exponential model on the left shows that there was a turning point at $\alpha = 0.5$ with the highest RMSE and the least RMSE is for values $\alpha = 0$ and $\alpha = 1.0$. For Holt's model the minimum RMSE values is for $\beta = 0$ and $\beta = 0.1$, and any other values increases the RMSE value.

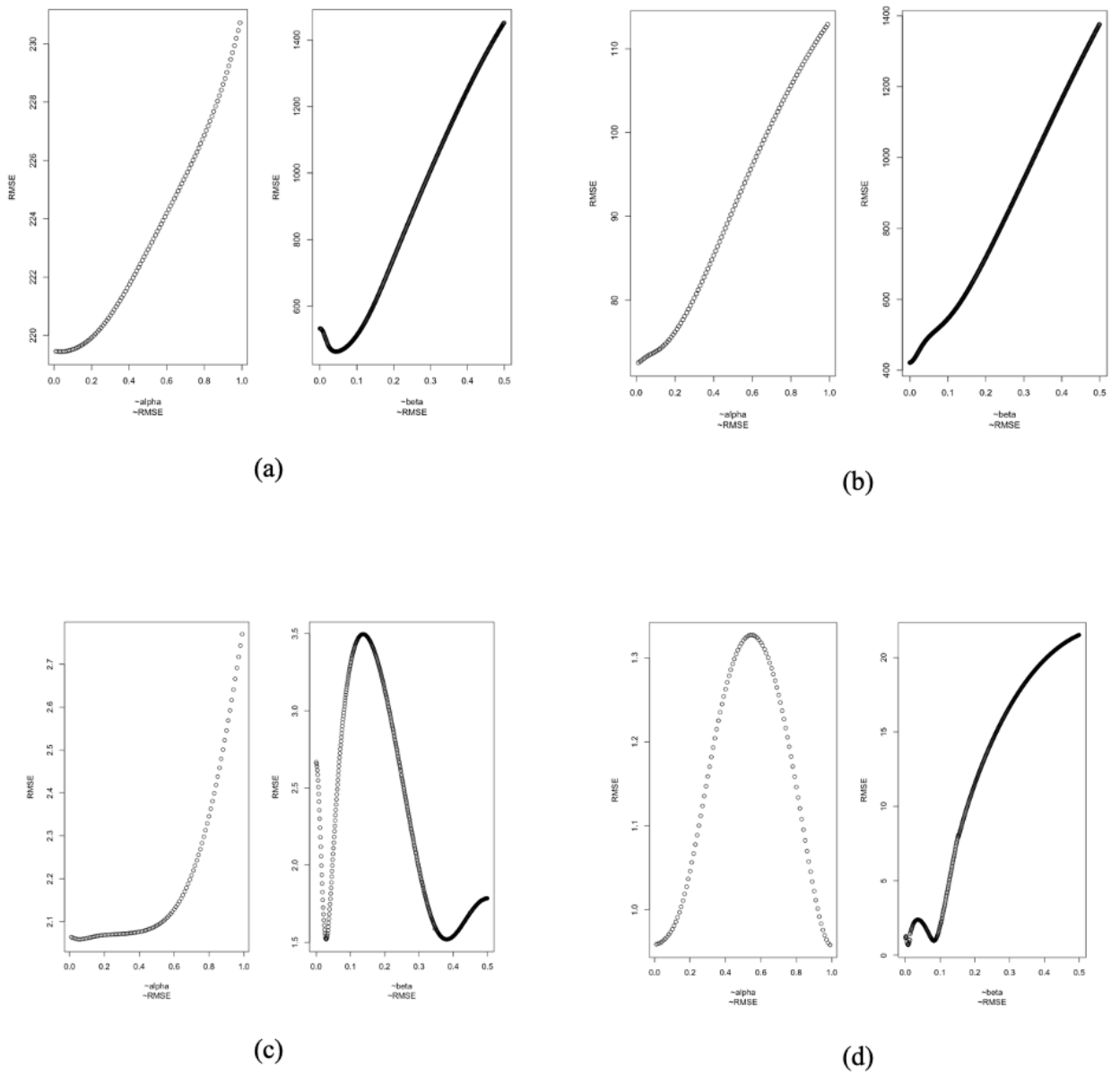


Figure 4.15: Comparison of simple exponential model without trend optimal RMSE and Holt's model optimal RMSE for (a) first wave cases, (b) second wave transition cases, (c) first wave deaths and (d) second wave transition deaths in Kuwait.

4.1.6 Sensitivity analysis

Here, we present the sensitivity analysis of the parameters used in the modelling of our study. The support vector model performance is sensitive to the choice of the cost function, γ and ϵ parameters for the data set and that is why we used the idea of cross validation and optimal model to choose the best parameters that best fit the model.
From Table 4.1, we present the sensitivity parameters for the cases we considered. We discovered that for ARIMA models, all the cases are more sensitive to the choice of the order of the auto regression and moving average while it is least sensitive to the trend difference.
For exponential models, the data is very sensitive to the choice of σ and α but least sensitive to the number of lags and choices of β. The optimal algorithm was able to choose the best

parameter that fit the model.

Table 4.1: The sensitive parameters for the analysis showing their Root MSE results (Note: FWC means first wave new cases, SWC means transition to second wave new cases, FWD means first wave deaths and SWD means transition to second wave deaths).

FWC	**SWC**	**FWD**	**SWD**
ARIMA(6,1,0) = 281.5	ARIMA(6,1,0) = 200.7	ARIMA(6,1,0) = 3.0	ARIMA(6,1,0) = 2.0
ARIMA(2,1,0) = 98.1	ARIMA(0,1,2) = 73.1	ARIMA(2,1,3) = 1.9	ARIMA(0,1,1) = 1.6
ARIMA(0,1,1) = 30.7	Holt's = 82.0	ARIMA(1,1,1) = 2.4	ARIMA(0,1,0) = 0.9
Holt's = 30.7	SES = 76.8	Holt's = 2.3	MNN = 1.7
SES = 31.4	diff.SES = 90.7	SES = 2.3	Holt's = 1.9
MAN = 111.0	-	MAN = 2.2	SES = 1.9
diff.SES = 38.3	ARIMA(0,1,2) with drift = 74.5	diff.SES = 2.7	diff.SES = 2.5

4.1.7 Prophet and neural prophet forecasting

Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends fit yearly, weekly, and daily data seasonal and holiday effects while neural prophet is a auto-regressive deep learning tool use for forecasting by using classical neural networks components. We observed that the death data from Kuwait has a better performance than the new case data. We present the visualisation and 100 days forecasting of the results in Figure 4.16(a)-(d) and the trend plots in Figure 4.17(a)-(d). Also we present in Table 4.3 and Table 4.2 the predicted values for Kuwait daily new cases and deaths for October, 2020 along with predicted range. It is observed that some of the numerical values generated using this model are close to the observed values of the COVID-19 pandemic in Kuwait. The mean square error is given as follows: first wave cases = 79247.71, second wave transition cases = 40288.81. The MSE for first wave deaths = 9.04 and second wave transition deaths = 3.93.

We have also presented in Figure 4.17(a)-(d) a loss plot for our neural network model which helps to know the performance of the model. It was observed that the loss plot shows a good fit and convergence of the model. The plot of training loss decreases to a point of stability and also the plot of validation loss decreases to a point of stability and has a small gap (generalisation gap) with the training loss. In Figure 4.17e-f we provide the visualisation of the neural forecast of 100 days for the first wave cases and second wave transition cases which aligns with the trend of results we have presented in other forecasting results in the previous Section, and also from the observed results from worldometer, we noted a decrease in daily cases from end of April to May 2021 and this also aligns with the result presented in Table 4.2.

The two phases of the COVID-19 outbreak in Kuwait present differences: (a) the first wave start shows an increase of the daily new cases (with a peak of 1000 daily new cases) during about 60 days with a slope of the exponential regression equal to 0.07, followed by an endemic phase with a high mean number of daily new cases (about 600) and with a delay of about 15 days an increase of deaths and (b) the second wave transition shows a decrease of the new cases during about 50 days with a slope of the exponential regression equal to -0.009, followed by an increase during about 20 days.

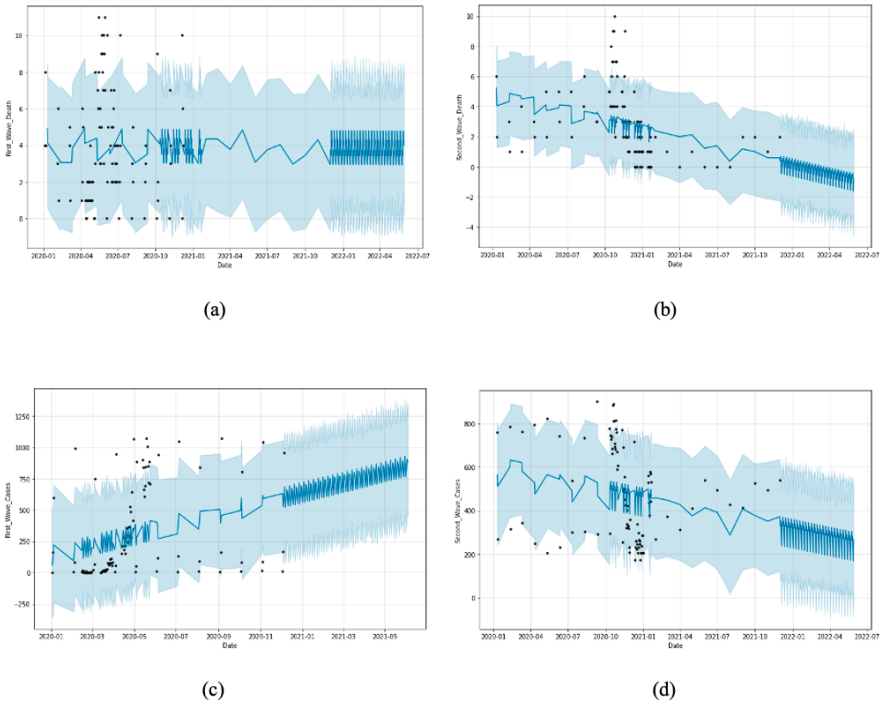


Figure 4.16: Prophet forecast for (a) first wave deaths, (b) second wave transition deaths, (c) first wave cases and (d) second wave transition cases in Kuwait.

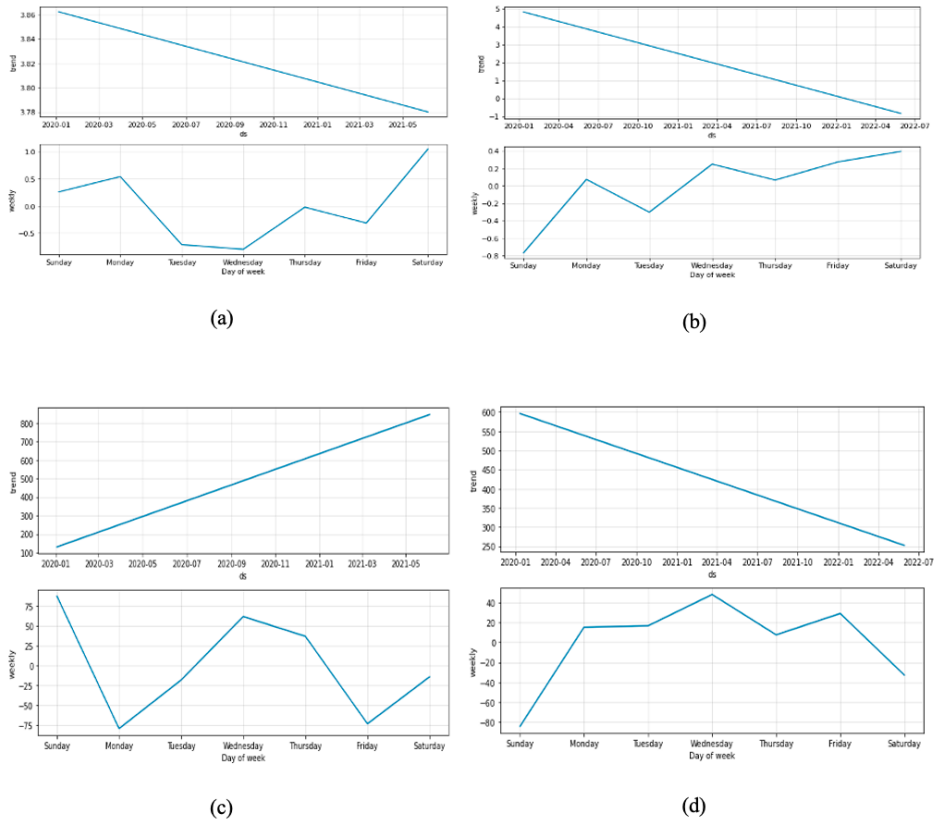


Figure 4.17: Trend plots for (a) first wave deaths, (b) second wave transition deaths, (c) first wave cases and (d) second wave transition cases in Kuwait.

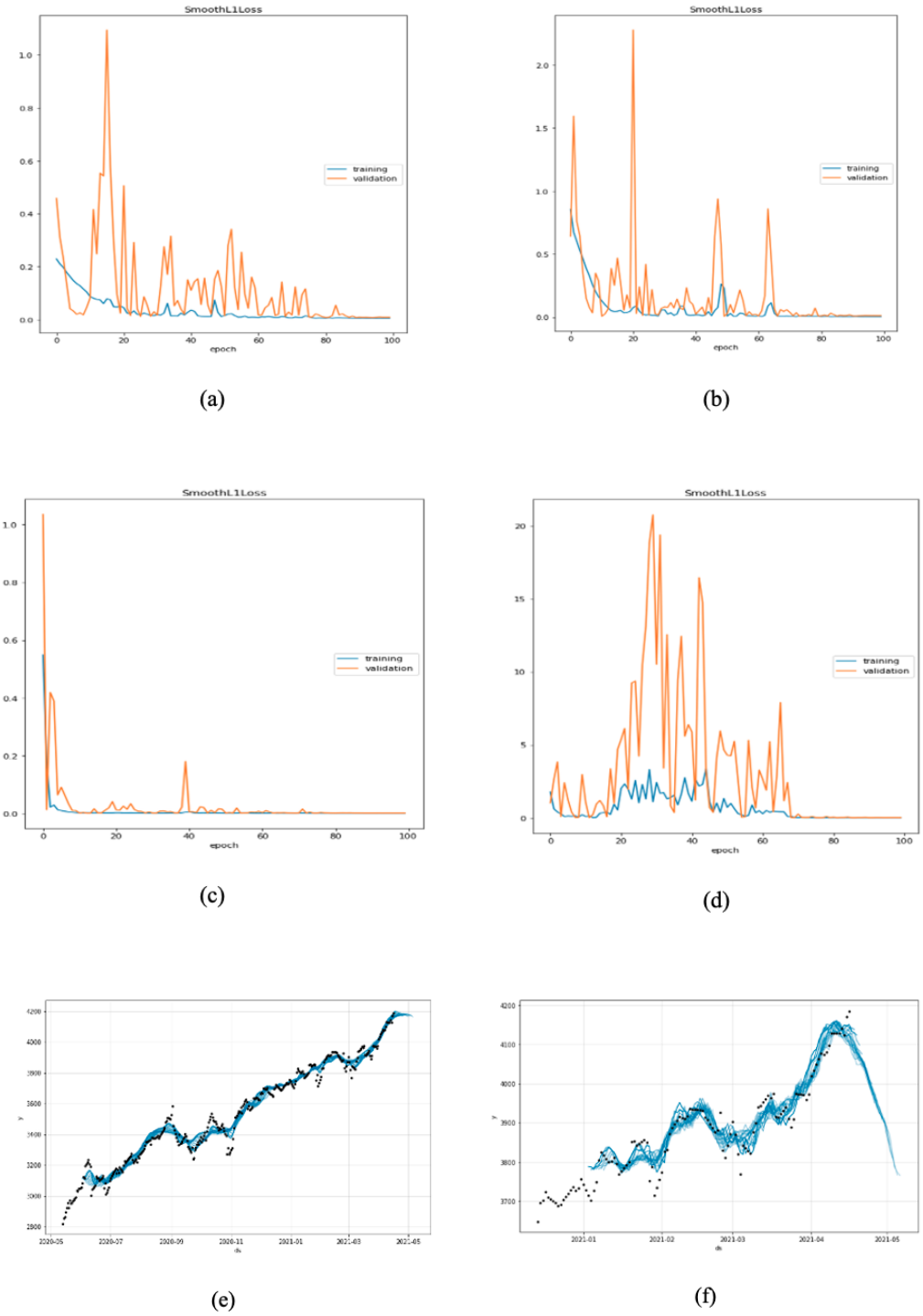


Figure 4.18: (a) Loss plot for second wave transition deaths in Kuwait, (b) Loss plot for second wave transition cases in Kuwait, (c) Loss plot for first wave deaths in Kuwait, (d) Loss plot for first wave new cases in Kuwait, (e) Neural forecast for first wave cases in Kuwait and (f) Neural forecast for second wave transition cases in Kuwait.

Table 4.2: Kuwait number of daily deaths and daily cases predicted for April and May, 2021 (Note: PRD means predicted range for daily deaths, PVD means predicted values for daily deaths, PRC predicted range for daily cases and PVC means predicted value for daily cases).

No	Date	PRD	PVD	PRC	PVC
1	17-04-21	1 ∼ 8	5	344 ∼ 1223	766
2	18-04-21	1 ∼ 8	4	467 ∼ 1303	869
3	19-04-21	1 ∼ 8	3	267 ∼ 1164	704
4	22-04-21	0 ∼ 7	4	380 ∼ 1276	824
5	23-04-21	0 ∼ 7	3	237 ∼ 1172	716
6	24-04-21	1 ∼ 9	5	345 ∼ 1241	776
7	25-04-21	0 ∼ 8	4	445 ∼ 1303	879
8	26-04-21	1 ∼ 8	4	268 ∼ 1131	714
9	27-04-21	0 ∼ 7	3	336 ∼ 1233	776
10	28-04-21	0 ∼ 7	3	397 ∼ 1280	857
11	29-04-21	0 ∼ 8	4	395 ∼ 1271	834
12	30-04-21	0 ∼ 7	3	281 ∼ 1166	725
13	01-05-21	1∼ 8	5	340 ∼ 1216	786
14	02-05-21	0∼ 8	4	439 ∼ 1311	888
15	03-05-21	0 ∼ 8	4	282 ∼ 1151	723
16	04-05-21	0 ∼ 7	3	369 ∼ 1228	786
17	05-05-21	0 ∼ 7	3	455 ∼ 1313	867
18	06-05-21	0 ∼ 8	4	387 ∼ 1274	844
19	07-05-21	0 ∼ 7	3	280 ∼ 1178	735
20	08-05-21	1 ∼ 8	5	357 ∼ 1215	795
21	09-05-21	0∼ 8	4	478 ∼ 1342	898
22	10-05-21	1 ∼ 8	4	291 ∼ 1155	733
23	11-05-21	0 ∼ 7	3	371 ∼ 1230	795
24	12-05-21	0 ∼ 6	3	457 ∼ 1345	877
25	13-05-21	0 ∼ 8	4	447 ∼ 1302	853
26	14-05-21	0 ∼ 7	3	330∼ 1211	744
27	15-05-21	1 ∼ 9	5	383 ∼ 1259	805
28	16-05-21	0 ∼ 8	4	448 ∼ 1369	908
29	17-05-21	1 ∼ 8	4	308 ∼ 1193	743
30	18-05-21	0 ∼ 7	3	381 ∼ 1245	805
31	19-05-21	0 ∼ 7	3	450 ∼ 1333	886
32	20-05-21	0∼ 7	4	441 ∼ 1291	863
33	21-05-21	0 ∼ 7	3	299 ∼ 1184	754
34	22-05-21	1 ∼ 9	5	332 ∼ 1284	815
35	23-05-21	0 ∼ 7	4	467 ∼ 1354	917
36	24-05-21	1 ∼ 8	4	279 ∼ 1184	753
37	25-05-21	0 ∼ 7	3	362 ∼ 1262	815
38	26-05-21	0 ∼ 7	3	494 ∼ 1332	896
39	27-05-21	0 ∼ 7	8	406 ∼ 1299	873
40	28-05-21	0 ∼ 7	3	340 ∼ 1213	764
41	29-05-21	1 ∼ 9	5	367 ∼ 1241	824
42	30-05-21	0 ∼ 7	4	510 ∼ 1375	927
43	31-05-21	1 ∼ 8	4	328 ∼ 1207	762

Table 4.3: Kuwait number of daily deaths predicted for October, 2020 (Note: PR means predicted range, OV means Observed value from worldometer and PV means predicted value).

Date	PR	OV	PV	In range?
11/10/20	$0 \sim 5$	3	2	YES
12/10/20	$0 \sim 6$	6	3	YES
15/10/20	$0 \sim 6$	8	3	NO
16/10/20	$0 \sim 6$	6	3	YES
17/10/20	$1 \sim 6$	4	3	YES
18/10/20	$0 \sim 5$	7	2	NO
19/10/20	$0 \sim 6$	9	3	NO
20/10/20	$0 \sim 5$	4	3	YES
21/10/20	$1 \sim 6$	7	3	NO
22/10/20	$0 \sim 6$	9	3	NO
23/10/20	$0 \sim 6$	10	3	NO
24/10/20	$1 \sim 5$	4	3	YES
25/10/20	$0 \sim 5$	2	2	YES
26/10/20	$0 \sim 6$	3	3	YES
27/10/20	$0 \sim 5$	7	3	NO
28/10/20	$0 \sim 6$	7	3	NO
29/10/20	$0 \sim 6$	4	3	YES
30/10/20	$0 \sim 6$	6	3	YES
31/10/20	$0 \sim 6$	6	3	YES

4.2 Approach to COVID-19 Time Series Data Using Deep Learning Method

This section aims to use a data-driven approach to retro-predict and forecast the daily new cases and deaths of the COVID-19 outbreak from the start of the pandemic until September 3, 2021 at various stages of the waves in USA, India, Brazil, France, Turkey, UK, and Russia, which are the top seven countries in the world where the pandemic is the most prevalent. For prediction, various deep learning methods was used, while for short term forecasting, ELM, MLP and spectral analysis were used, and comparison of their performance was done using root mean square error (RMSE) and relative root mean square error (rRMSE).

4.2.1 Time series and evaluation metrics

Here, the time series visualization (see Figure 4.19) for Turkey, the United Kingdom, the United States, Russia, India, France, and Brazil for daily new cases, 3-day moving average, and 7-day moving average from [119] is presented. This time series data from [119] will be used throughout the analysis.

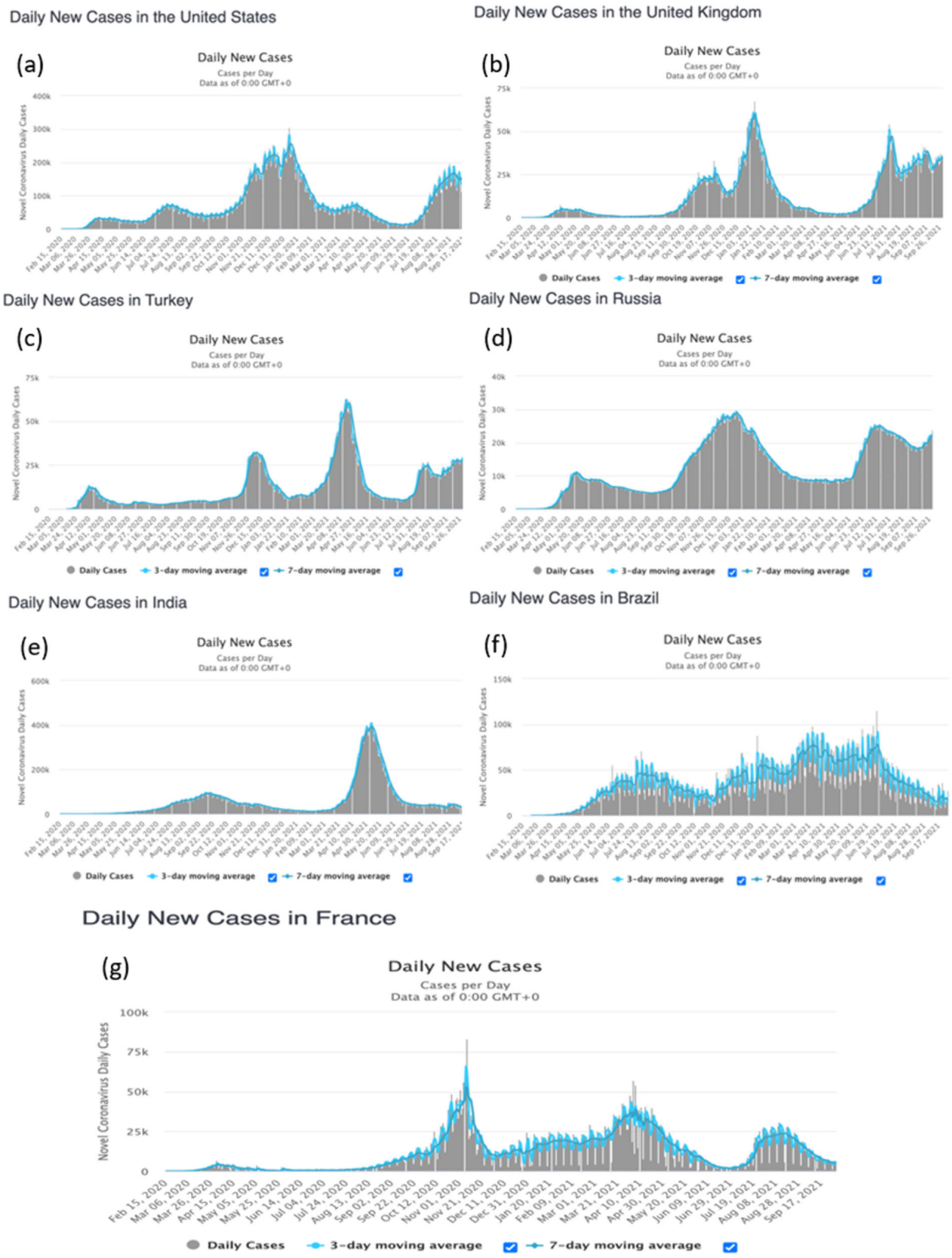


Figure 4.19: Times series visualization of daily new cases for (a) USA, (b) UK, (c) Turkey, (d) Russia, (e) India, (f) Brazil and (g) France.

When analyzing a model, it is critical to measure its performance in order to draw the best

conclusion and interpretation for the time series data. Two errors were used to estimate the prediction and forecasting precision for the models, where for $i = 1, 2, \ldots, n, Y_i$'s are the observed values, n is the number of data points and y_i's are the predicted values given below:
- Root Mean Square Error (RMSE), given as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - y_i)^2},$$

- relative Root Mean Square Error (rRMSE), given as:

$$\text{rRMSE} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - y_i)^2}}{\max(Y_i) - \min(Y_i)}.$$

4.2.2 Parameters for the modeling in LSTM, GRU, CNN, and DNN methods

To evaluate the model, 80% of the data set was trained and 20% was tested. The MinMaxScaler feature in Python was used to normalize the data. The errors are shown in Table 4.4 and Table 4.5. Figure 4.20 shows the visualization of the results for daily new cases and Figure 4.21 for daily deaths. Other specific details of the parameter used for each model is presented as follows:

- LSTM: The best hyperparameter tuning for the LSTM, as determined by a manual search, is: Batch size of 32, epochs of 100, drop out = 0.2, and units = 50. Adam optimizer was used, and mean square error measures the effectiveness of the loss. The Adam optimizer was found to slightly outperform other learning algorithms.

- GRU: the best hyperparameter tuning for the model, as determined by a manual search, is: Batch size of 30, epochs of 100, drop out = 0.2, and units = 50. Adam optimizer was used and tanh activation function. Mean square error measures the effectiveness of the loss. The Adam optimizer was found to slightly outperform other learning algorithms.

- CNN: An input layer with three neurons was found, followed by convolution layers with 128 units, 64 units, and 16 units, a max-pooling layer with 64 units, and a kernel size of 3. The best hyperparameter tuning for the model, as determined by a manual search, is: Batch size of 32, epochs of 100, verbose = 0.2 but 0 for checkpoint, and validation split = 0.2. Adam is the optimizer in use. While the mean square error was used to assess the effectiveness of the loss, the relu activation function was used for all layers expecting the output with linear function. Adam optimizer was found to slightly outperform other learning algorithms.

- DNN: The DNN has two hidden layers with 32 and 8 neurons, respectively, and it was discovered that if the model goes beyond 32 neurons, the model tends to give much higher error with poorer prediction. The best hyperparameter tuning for the model, as determined by a manual search, is: Batch size of 32, epochs of 100, verbose = 0.2 but 0 for checkpoint, period = 1, and validation split = 0.2. The optimizer used is Adam, and mean square error measures the effectiveness of the loss. The relu activation function was used for all layers expecting the output with linear function. Adam optimizer was found to slightly outperform other learning algorithms.

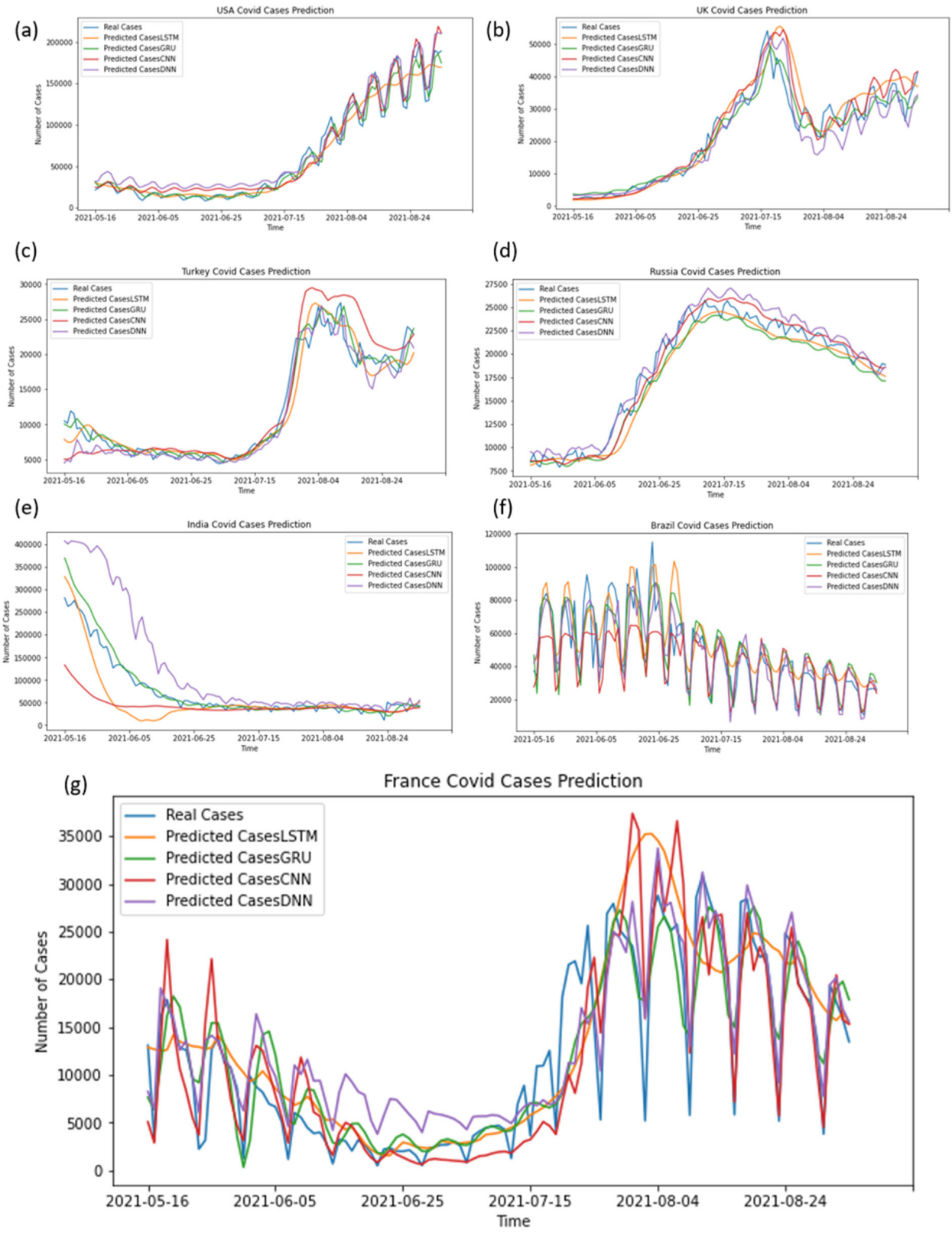


Figure 4.20: LSTM, GRU, CNN, and DNN results of daily new cases for (a) USA, (b) UK, (c) Turkey, (d) Russia, (e) India, (f) Brazil and (g) France.

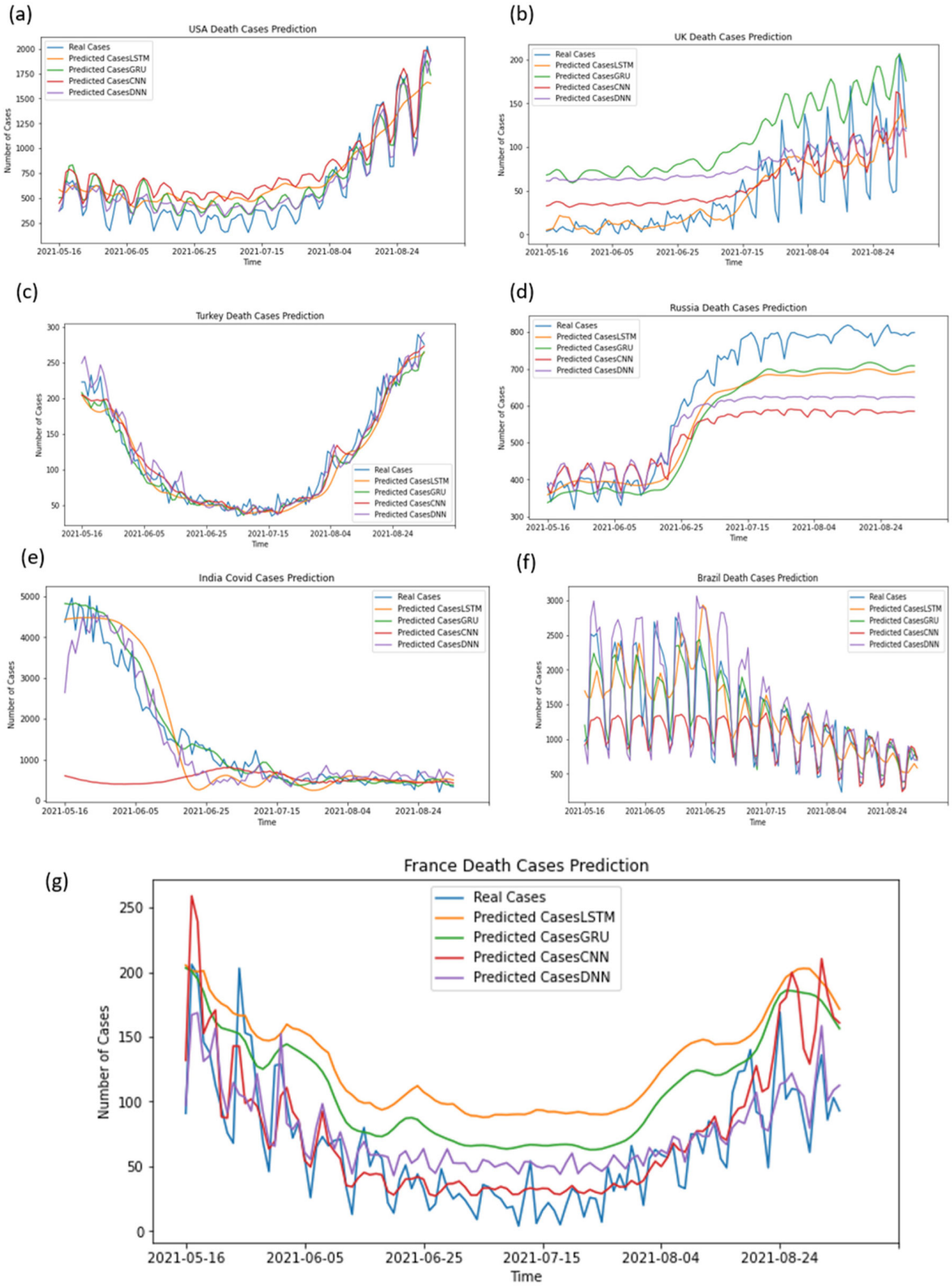


Figure 4.21: LSTM, GRU, CNN, and DNN results of daily deaths for (a) USA, (b) UK, (c) Turkey, (d) Russia, (e) India, (f) Brazil and (g) France.

4.2.3 Parameters for the modeling in ANN method

Here, a visualization of the network weights and the bias between the data is presented. As shown in Figure 4.22 and Figure 4.23, the weights are good with low bias. Also, some of the data were predicted (the test data) and the daily new cases prediction scores in Figure 4.22 for (a) France is 1.0, (b) India is 0.99, (c) USA is 1.00, (d) Brazil is 0.99, (e) Russia is 0.99, (f) Turkey is 0.10, and (g) UK is 0.99. In Figure 4.23, shows the daily deaths prediction scores for (a) France is 0.93, (b) India is 0.85, (c) USA is 0.99, (d) Brazil is 0.88, (e) Russia is 0.15, (f) Turkey is 0.99, and (g) UK is 0.51.
It was discovered that the daily deaths errors were lower than those of the daily new cases.

4.2.4 Parameters for the modeling in MLP and ELM methods

Here, the description of the best parameter used for each case based on performance and forecasting of daily new cases and deaths. For Turkey daily deaths and Russia daily new cases and deaths, 12 lags, five hidden layers, and 20 repetitions were used, with series modelled in difference and forecast combined using the median operator, as shown in Figure 4.24(d), Figure 4.25(c) and Figure 4.25(d).
For daily new cases, 24 lags for Turkey, India, and Brazil was used, as well as for death cases for the United Kingdom while keeping 12 lags and testing the rest for inclusion. Other parameters, such as those shown in Figure 4.24(c), Figure 4.24(f), Figure 4.25(b) and Figure 4.25(e), were kept. In addition, for daily new cases in France, United States, United Kingdom, and Brazil, 24 lags was used while retaining all lags, as shown in Figure 4.24(a), Figure 4.24(b), Figure 4.24(g) and Figure 4.25(f). For India daily death cases, four hidden layers was used, 20 repetitions, and 12 lags while keeping the same parameters for forecasting and series modelling as shown in Figure 4.24(e). Figure 4.25(a) shows the use of 24 lags for daily deaths in the United States, with no testing for inclusion. Finally, for France daily deaths, the trend with no differencing was removed while also using regressors and 12 lags. The errors are shown in Table 4.4 and Table 4.5.
According to Table 4.4 and Table 4.5, DNN has the least error for the daily new cases in the United States, and by looking at Figure 4.20(a), one can say that the estimation is good and it reflects the situation of the pandemic in the country as new cases fluctuate, just like daily deaths, which are best predicted by GRU with the least error. For the UK, MLP performed better than other models with the least error, as shown in Figure 4.24(b), where daily new cases are decreasing from November 2021 as observed from real data, while CNN best predicts daily deaths, as shown in Figure 4.21(b). For Russia, the daily new cases and daily deaths are best predicted by MLP, which has the least error, and from Figure 4.24(d) and Figure 4.25(d), we can see that the forecast curve is increasing, which corresponds to the current situation in the country, and with the festive season approaching in December 2021, it is expected that cases will increase if not controlled and also as new variants emerges because the virus keeps mutating. For France, CNN best predicts daily deaths, indicating that, despite a decreasing trend, cases will fluctuate as shown in Figure 4.21(g), and MLP best predicts daily new cases, indicating a slight increase in September 2021 as shown in Figure Figure 4.24(g). We discovered that MLP achieves the best prediction for daily new cases and daily deaths in India, Brazil, and Turkey, as shown in Figure 4.24c, Figure 4.24(e), Figure 4.24(f), Figure 4.25(c), Figure 4.25(e), and Figure 4.25(f). Despite a slight increase in September cases, the pandemic dynamics from October in India, Brazil, and Turkey shows a decreasing but fluctuating trend.

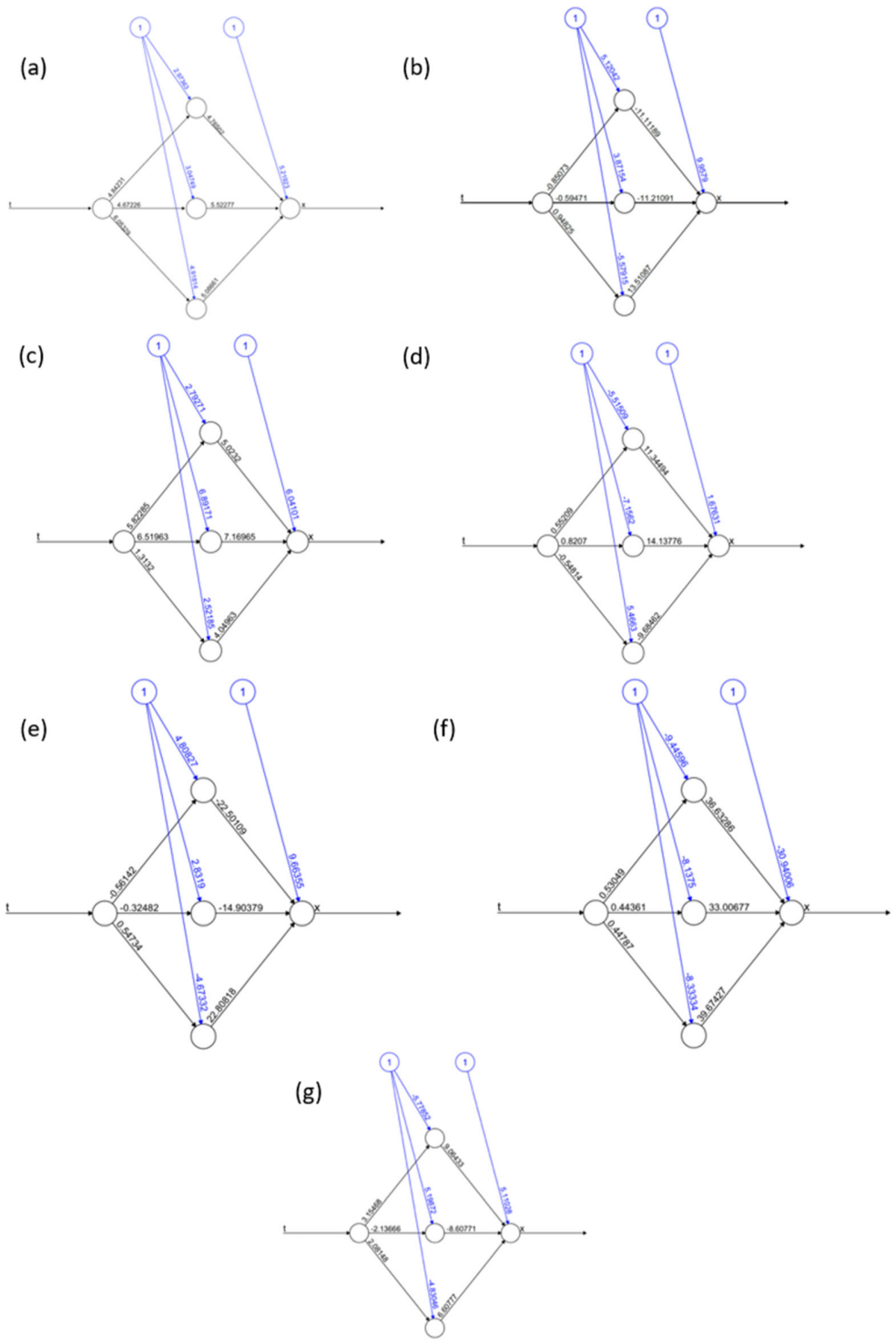


Figure 4.22: Neural network for daily new cases in (a) France, (b) India, (c) USA, (d) Brazil (e) Russia, (f) Turkey, and (g) UK. In figures, x represents daily new cases and t represents days.

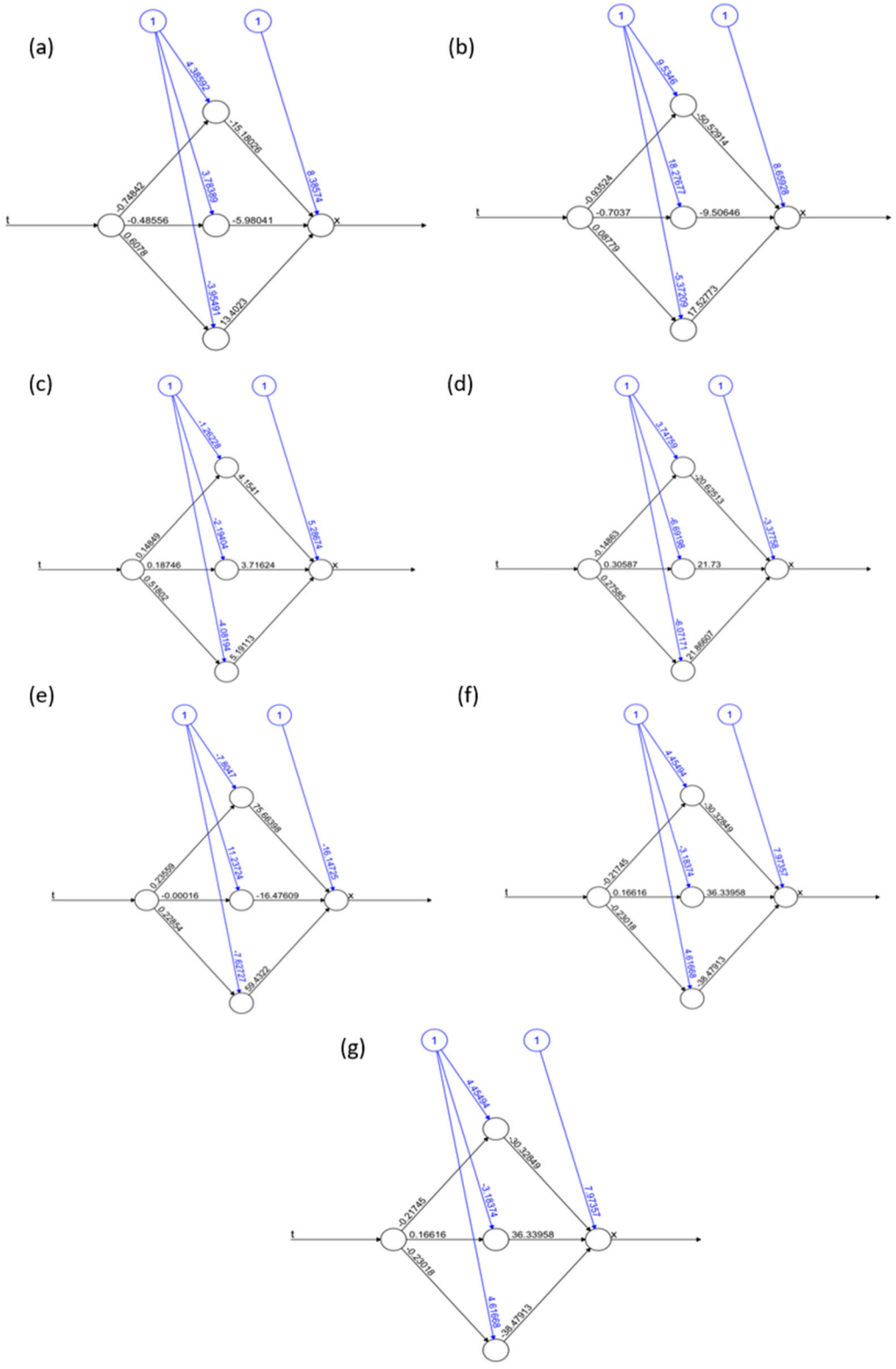
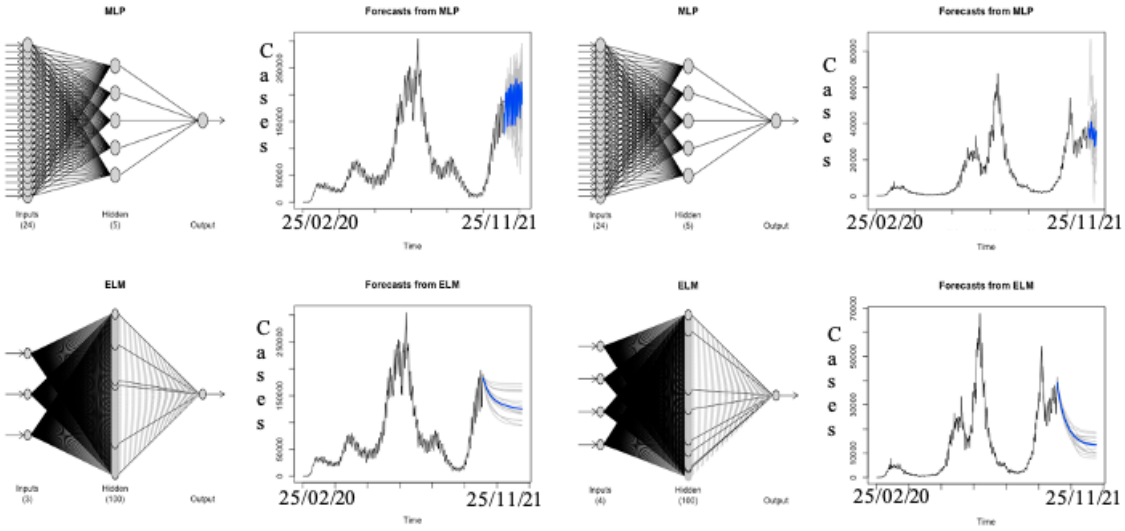


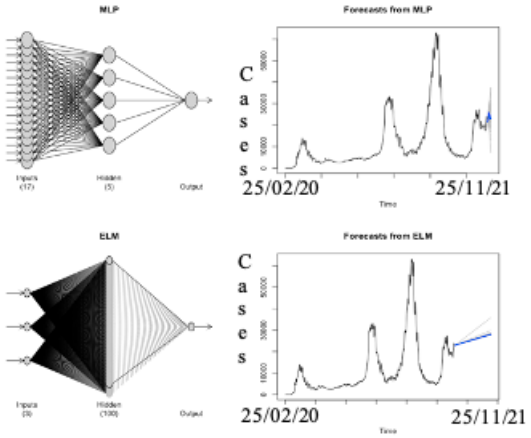
Figure 4.23: Neural network for daily deaths in (a) France, (b) India, (c) USA, (d) Brazil (e) Russia, (f) Turkey, and (g) UK. In figures, x represents daily deaths while t represents days.


MLP
Forecasts from MLP
ELM
Forecasts from ELM
Cases
25/02/20
25/11/21
Time

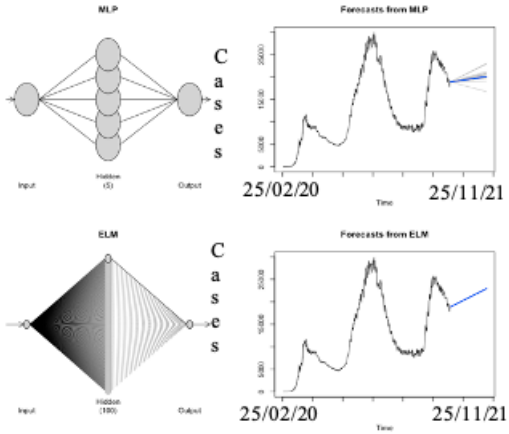
(a)

MLP
Forecasts from MLP
ELM
Forecasts from ELM
Cases
25/02/20
25/11/21
Time

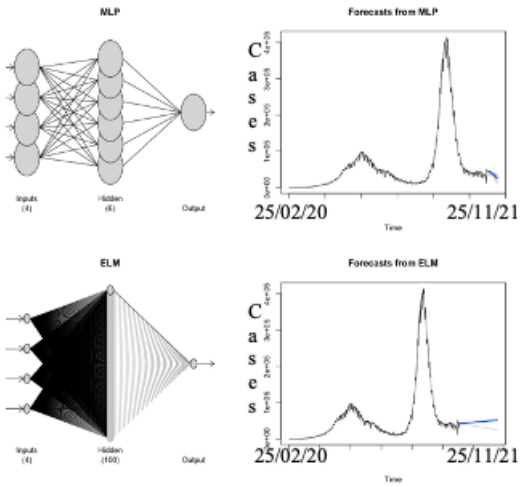
(b)


MLP
Forecasts from MLP
ELM
Forecasts from ELM
Cases
25/02/20
25/11/21
Time

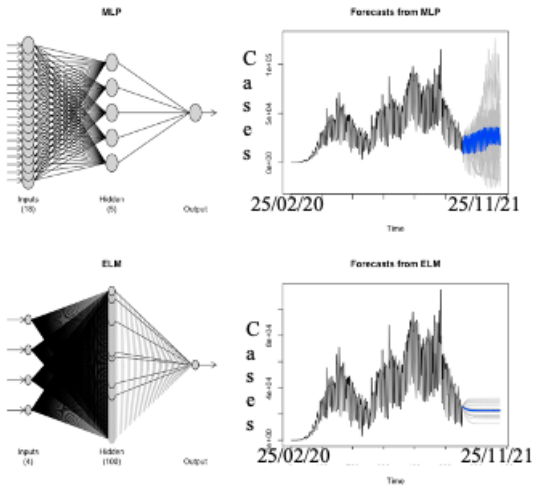
(c)


MLP
Forecasts from MLP
ELM
Forecasts from ELM
Cases
25/02/20
25/11/21
Time

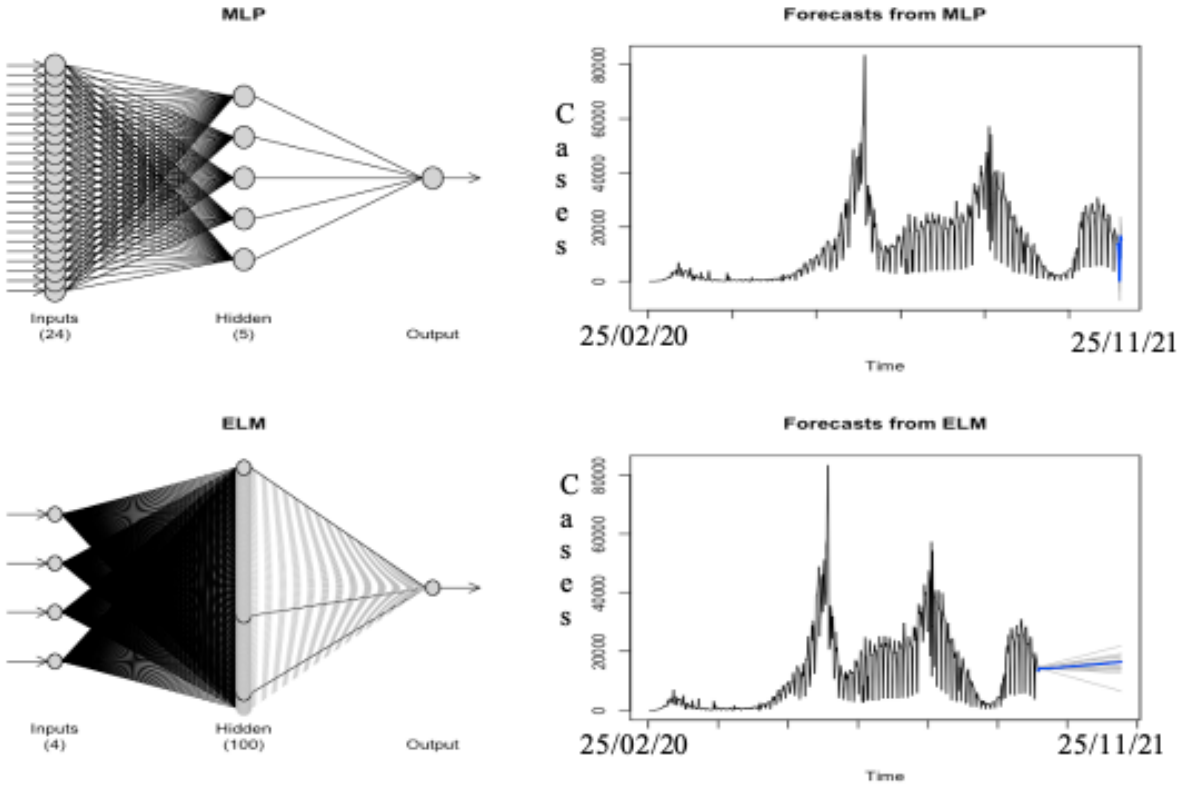
(d)


MLP
Forecasts from MLP
ELM
Forecasts from ELM
Cases
25/02/20
25/11/21
Time

(e)


MLP
Forecasts from MLP
ELM
Forecasts from ELM
Cases
25/02/20
25/11/21
Time

(f)



(g)

Figure 4.24: MLP and ELM results of daily new cases for (a) USA, (b) UK, (c) Turkey, (d) Russia, (e) India, (f) Brazil and (g) France.

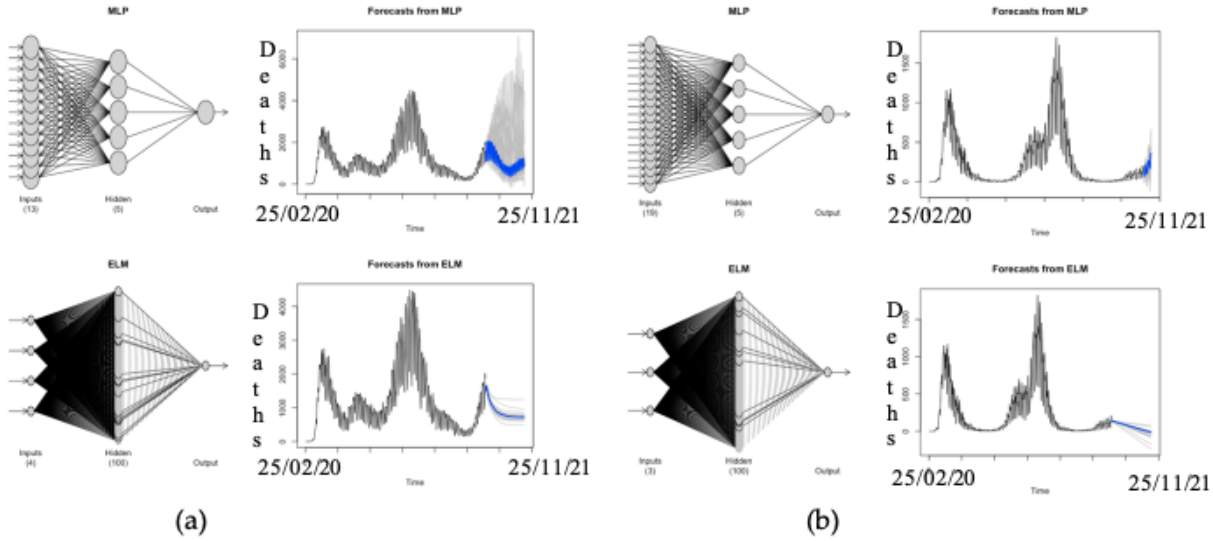
It is also observed that all the methods are sensitive to the parameters being used: for instance, for DNN if we go beyond 32 neurons, the model tends to give much higher error with poorer prediction. Also, for MLP if we use too many lags and neurons for some of the data, the prediction will be poorer, which was explained earlier in Section 4.2.2. The training time used in the modelling varies across models and that is one factor that enhances their performance: it is observed that LSTM training time was the lowest, hence, it does not perform better in any of the data set. Eventually, it was noticed also that the model performance depends highly on the data size.

Table 4.4: Comparison of the prediction performances of daily new cases for deep learning results.

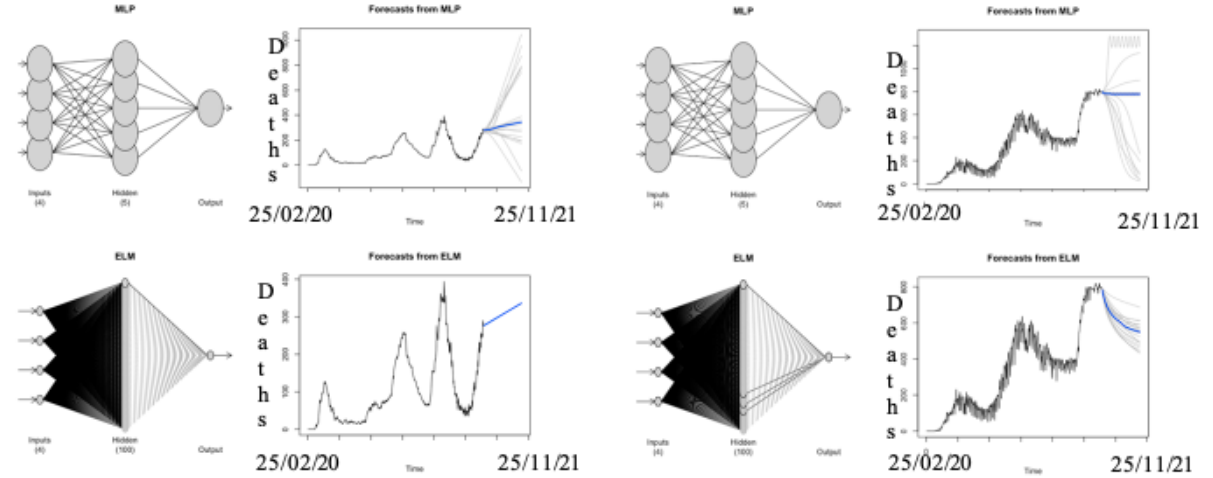
Country name	**RMSE LSTM**	**RMSE GRU**	**RMSE CNN**	**RMSE DNN**	**RMSE MLP**	**RMSE ELM**	**Mean daily cases**
USA	15484.45	10401.13	10308.60	9630.55	5031.00	11149.23	73122.00
India	28990.34	11526.90	76877.87	114120.26	7503.58	7956.88	59147.00
Brazil	16355.71	11223.93	15638.27	214359.20	5324.19	11700.93	37607.00
Turkey	2327.30	1577.54	2151.18	3936.00	578.38	1434.25	11595.00
France	6708.81	4734.50	4358.47	4062.04	1864.17	7163.29	12231.00
Russia	1728.23	1676.55	1165.26	1115.78	640.81	643.54	12523.00
UK	5022.61	4381.44	4205.14	5613.93	1330.47	2689.11	12394.00
	rRMSE LSTM	**rRMSE GRU**	**rRMSE CNN**	**rRMSE DNN**	**rRMSE MLP**	**rRMSE ELM**	
USA	0.08	0.06	0.05	0.05	0.11	0.11	
India	0.11	0.04	0.28	0.42	0.02	0.22	
Brazil	0.16	0.11	0.15	0.14	0.11	0.14	
Turkey	0.10	0.07	0.09	0.17	0.03	0.07	
France	0.22	0.16	0.14	0.13	0.10	0.22	
Russia	0.10	0.09	0.07	0.06	0.03	0.23	
UK	0.11	0.08	0.08	0.11	0.07	0.10	

Table 4.5: Comparison of the prediction performances of daily deaths for deep learning results.

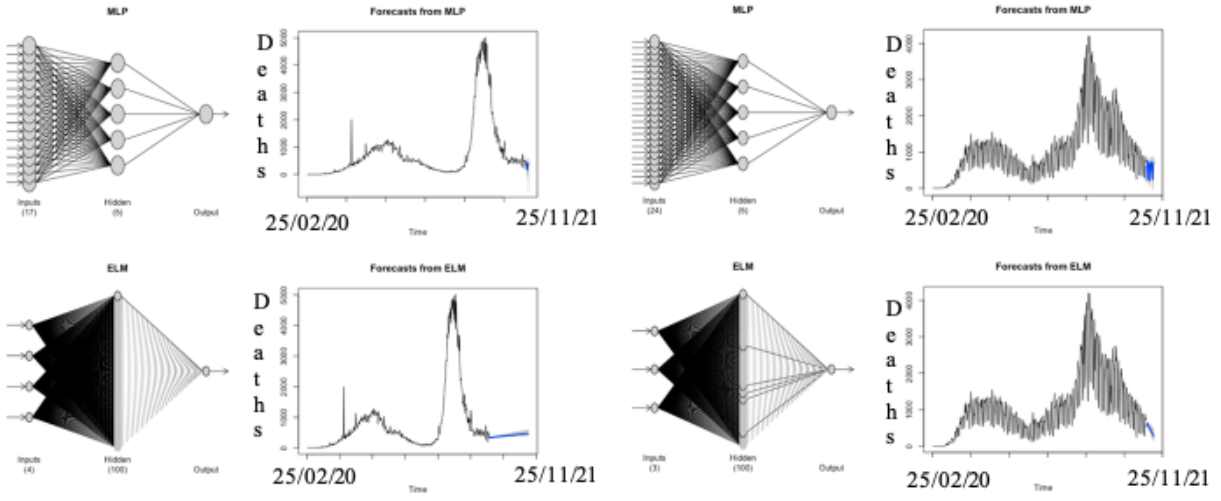
Country name	**RMSE LSTM**	**RMSE GRU**	**RMSE CNN**	**RMSE DNN**	**RMSE MLP**	**RMSE ELM**	**Mean daily deaths**
USA	225.66	105.15	249.55	194.62	130.77	288.84	1202.00
India	387.64	223.58	1670.94	301.49	95.26	182.50	790.40
Brazil	739.86	276.26	579.29	397.66	110.27	369.81	1046.00
Turkey	16.49	14.54	15.96	15.61	8.54	9.98	103.30
France	35.08	37.80	28.72	31.20	79.30	122.88	206.10
Russia	97.66	73.74	149.56	122.06	32.66	38.03	333.20
UK	29.43	37.82	27.32	69.16	35.77	124.91	238.90
	rRMSE LSTM	**rRMSE GRU**	**rRMSE CNN**	**rRMSE DNN**	**rRMSE MLP**	**rRMSE ELM**	
USA	0.12	0.06	0.13	0.10	0.24	0.13	
India	0.08	0.05	0.35	0.06	0.01	0.10	
Brazil	0.29	0.11	0.23	0.50	0.06	0.08	
Turkey	0.06	0.06	0.06	0.06	0.04	0.06	
France	0.17	0.19	0.14	0.15	0.15	0.24	
Russia	0.19	0.15	0.30	0.24	0.11	0.18	
UK	0.14	0.18	0.13	0.33	0.17	0.24	


MLP
Forecasts from MLP
Deaths
25/02/20
25/11/21
Time
ELM
Forecasts from ELM
Deaths
25/02/20
25/11/21
Time

(a) (b)


MLP
Forecasts from MLP
Deaths
25/02/20
25/11/21
Time
ELM
Forecasts from ELM
Deaths
25/02/20
25/11/21
Time

(c) (d)


MLP
Forecasts from MLP
Deaths
25/02/20
25/11/21
Time
ELM
Forecasts from ELM
Deaths
25/02/20
25/11/21
Time

(e) (f)

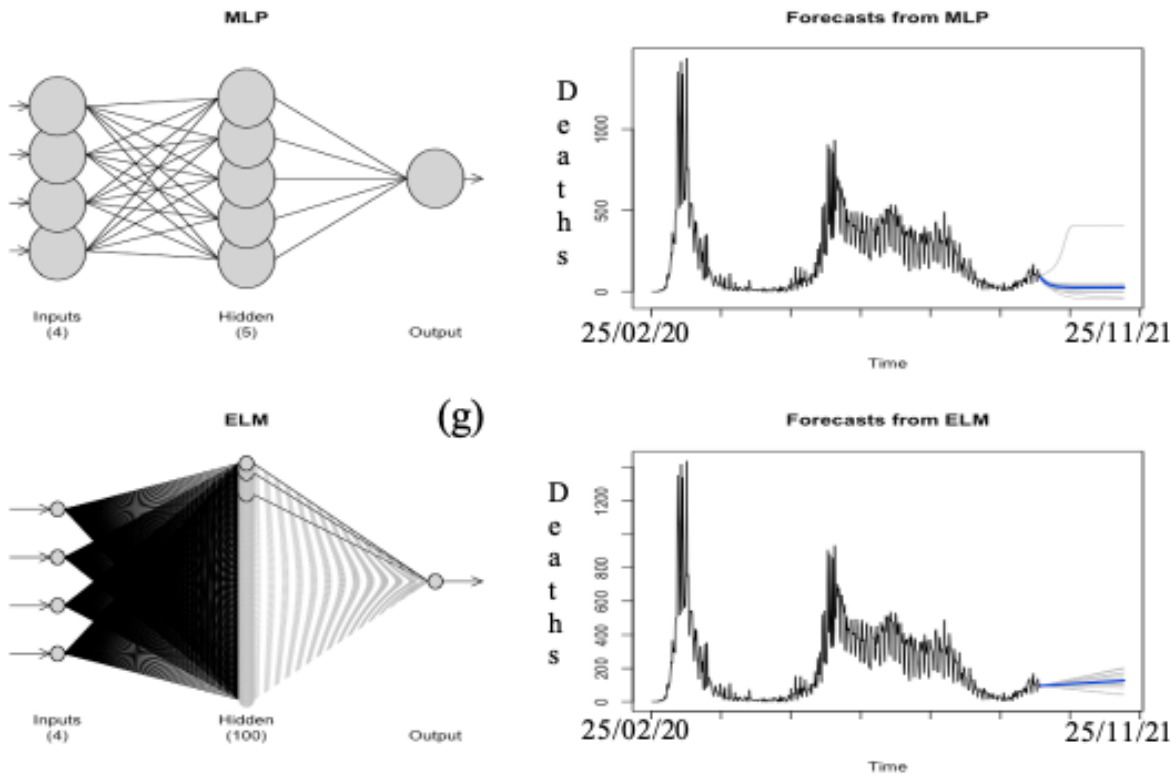


Figure 4.25: MLP and ELM results of daily deaths for (a) USA, (b) UK, (c) Turkey, (d) Russia, (e) India, (f) Brazil, (g) France.

4.3 Approach to COVID-19 Time Series Data Using Spectral Analysis Method

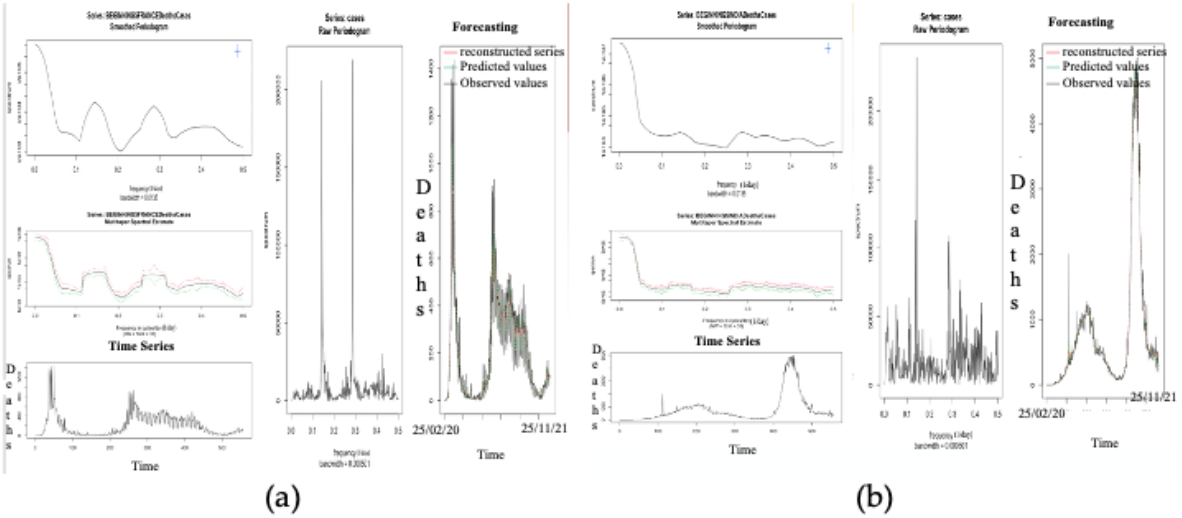
In this section, spectral analysis is used to measure the different frequencies in COVID-19 data set presented in Section 4.2.1 during the contagiousness period by estimating the time series periodicity and analyzing their peaks. Spectral analysis allows to remove the weekly and seasonal components discovered on the epidemiological data of incidence and mortality. Firstly, the stationarity and normality is checked before applying spectral analysis to the set of time series data. The Box-Lung test, the Jarque-Bera normality test, and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test was used and it was found that the p-value ≤ 0.01 for all of these tests and for all of the time series considered in this article, leading to the conclusion that the data set is stationary and its residual is normal. So, the estimation of the spectral density and smoothed the periodogram was done and also the plotting of the spectral density and spectrum and then performed harmonic regression because the goal of spectral analysis is to decompose a time series into periodic components. This was considered by performing a regression, in which the time series was regressed on a set of sine and cosine waves and then attempted to regress the time series on harmonic waves, including daily harmonics as well as other harmonics. For the entire data set, a p-value ≤ 0.05 is obtained. The daily variation were accurately described and the median of the residual was negative and close to zero, confirming the normality. The scaling of both x and y using a continuous scale for x

and log scale for y is done. Log-scaling has some theoretical advantages. The periodogram values should be roughly normally distributed in the log scale, and log scaling can be useful because it spreads out the low frequencies while squashing the high frequencies. Due to the noisy nature of the periodogram, it was smoothed by using 9 moving averages for the kernel functions of the periodogram. To estimate the spectral density, tapering and multi-taper was used, which is a method that allows to test for peaks using the F-test. It was discovered that p-value ≤ 0.001 for the entire data set, indicating that there are peaks in the model. The time bandwidth parameter (NW) used is 16 and the number of tapers is 31. Figure 4.27 and Figure 4.26 show the results of the spectral analysis as well as a forecasting using the analysis, methods and theoretical study given in [226].

Figure 4.27 and Figure 4.26 show a clear peak of frequency at value 0.145, which is approximately a 7 days period showing a weekly pattern for countries considered, both for daily new cases and daily deaths, except for Turkey daily deaths as seen in Figure 4.26(f), where peaks are not really visible with a slight peak at 0.14 and another at 0.08.

India daily deaths, Russia daily new cases, and the United Kingdom daily new cases have only one clear peak (see Figure 4.27(e), Figure 4.27(g) and Figure 4.26(b)), whereas others have smaller peaks at 3.6 days and 3.4 days, which represent approximately mid-week, which is an expected phenomenon due to accumulated cases from the beginning of the week, and as also seen in some countries collating their infection records after three days or more (e.g., one week in Cameroon).

Furthermore, the forecasting pattern of the spectral analysis results corresponds to that of the MLP forecasting results presented in Section 4.2.4, with the exception of Russia daily new cases and daily deaths, which have different results for both methods. Forecasting based on spectral analysis also confirms fluctuations in MLP forecasted values presented in Section 4.2.4.



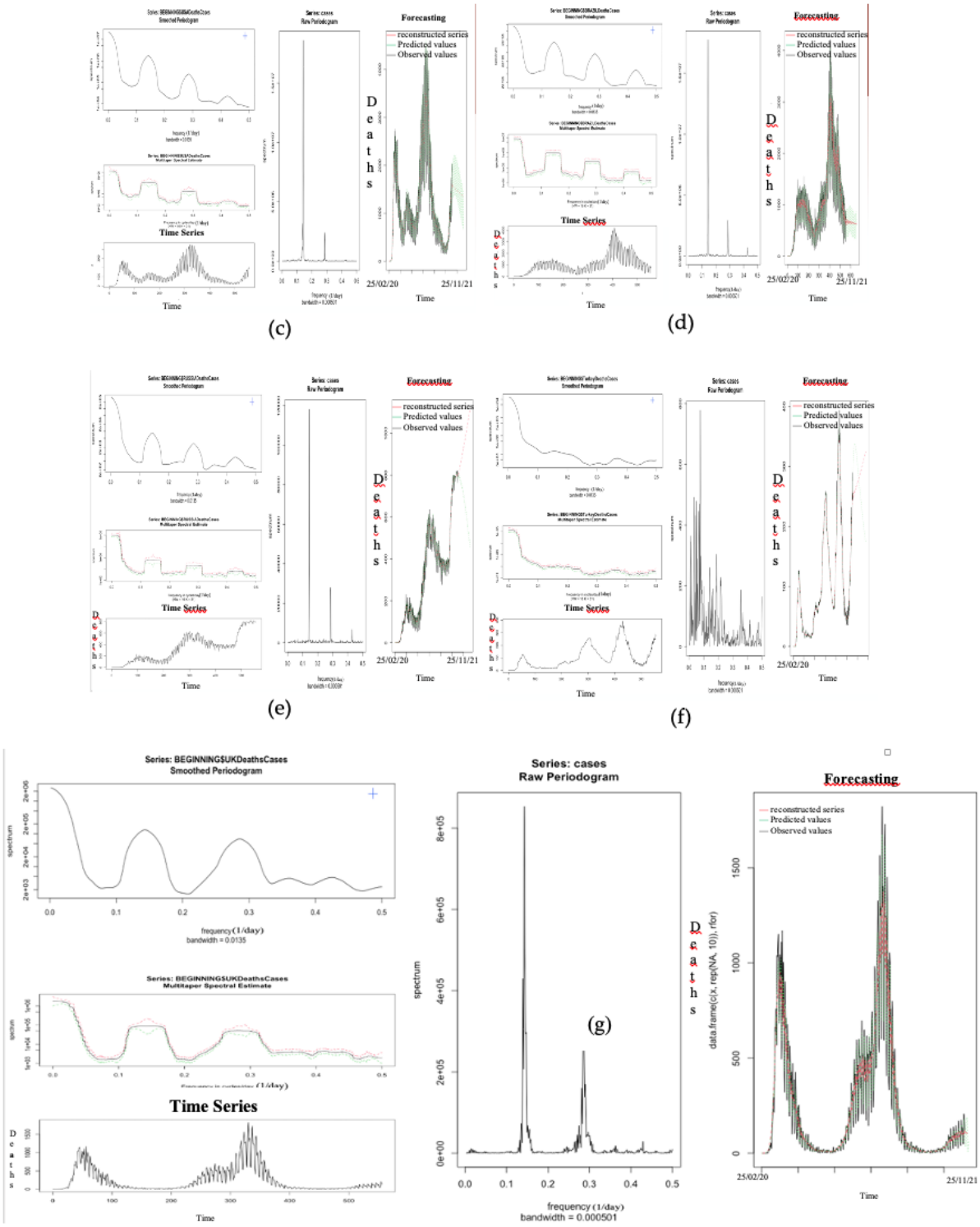
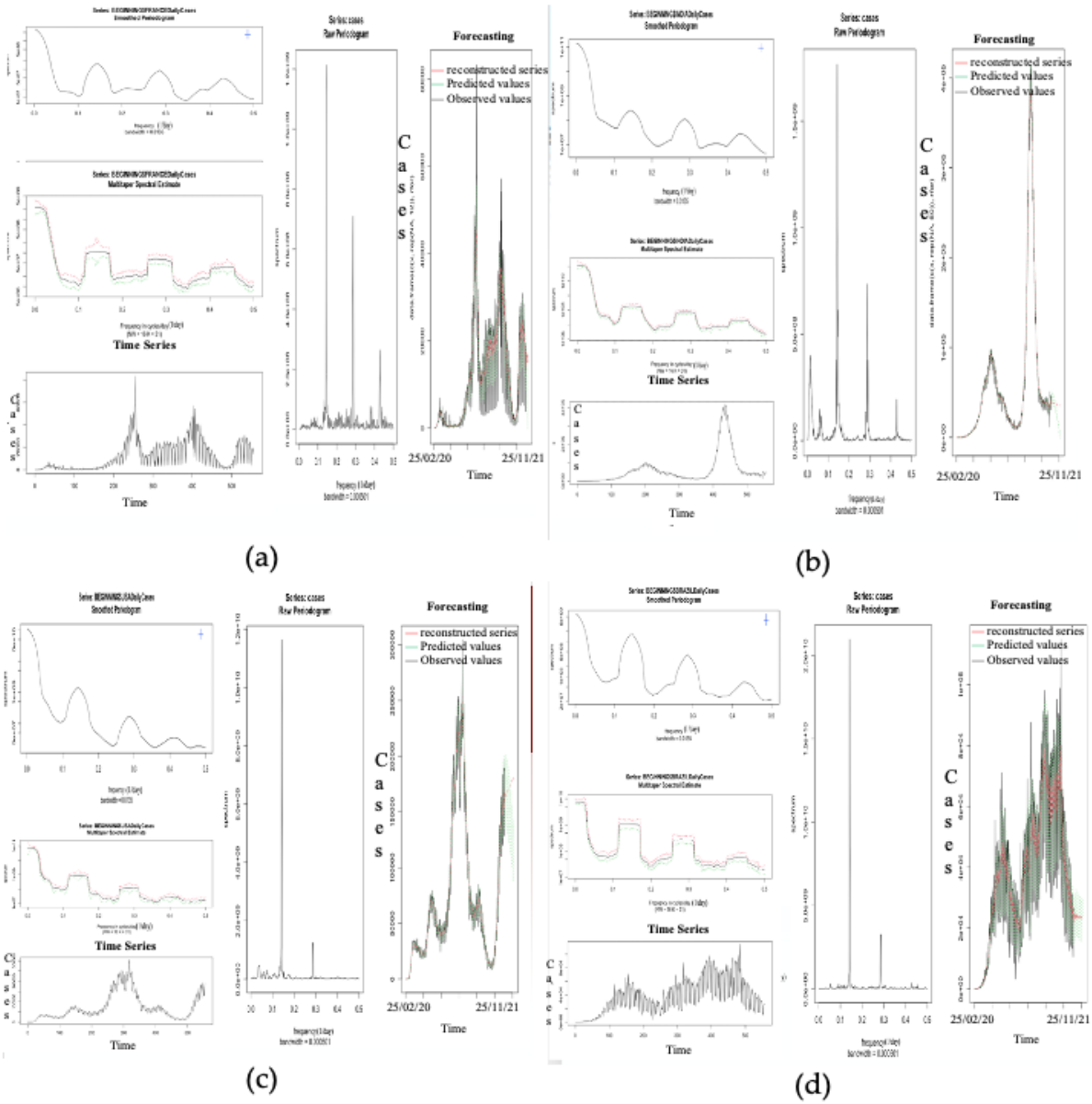


Figure 4.26: Spectral analysis and forecasting of daily deaths for (a) France, (b) India, (c) USA, (d) Brazil (e) Russia, (f) Turkey and (g) UK.


Time Series
Series: cases
Raw Periodogram
Forecasting
reconstructed series
Predicted values
Observed values
Cases
Time
25/02/20
25/11/21
(a)
(b)
(c)
(d)

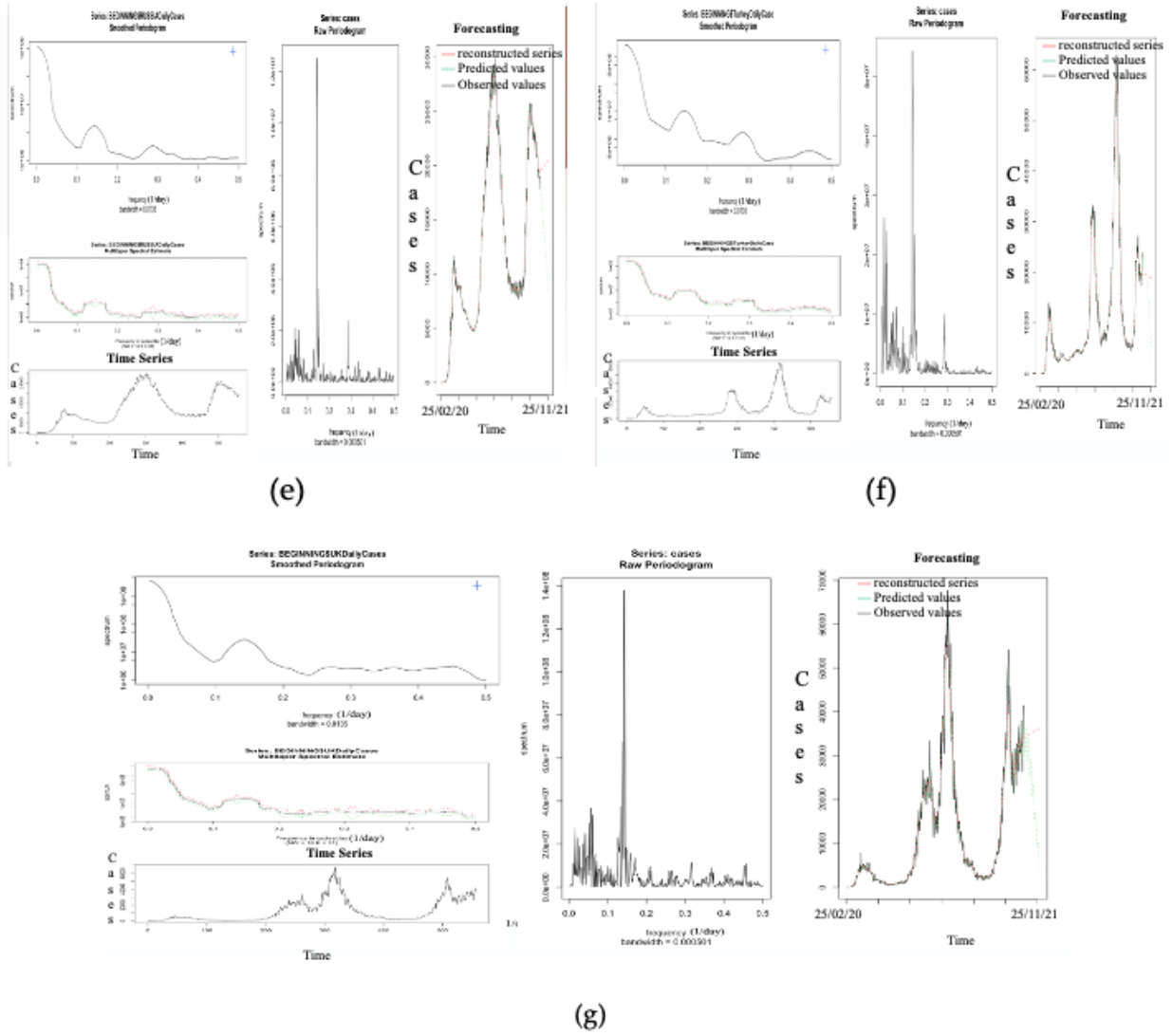


Figure 4.27: Spectral analysis and forecasting for daily new cases in (a) France, (b) India, (c) USA, (d) Brazil (e) Russia, (f) Turkey and (g) UK.

4.4 Modeling of COVID-19 Pandemic vis-à-vis Some Epidemiology, Socio-economic, Geo-climatic and Demographic Factors

The use of machine learning methods to analyze data has been helpful over the years to get a proper view on how a model behaves. In this Section, some supervised and unsupervised machine learning methods were used and also tried to use two deep learning methods. To jointly interpret the socio-economic, geo-climatic, demographic and epidemiological data, these main classes of the descriptive statistics was chosen, which allows to compare them. Supervised learning is used in its regression function (prediction of a quantitative variable from annotated examples) and unsupervised learning (in which the data is not labeled) in its classification function. As for deep learning, it makes it possible to create a model from

large-scale unlabeled data.

The supervised machine learning methods used are first univariate polynomial regression, linear regression, lasso regression, and ridge regression. Also, some of these methods was used to make prediction by training the model and testing some percentage of the values. Lasso regression helps to know the best variables to be used in the modeling. After the univariate regressions, multivariate least square methods was introduced, allowing to test much more complex relations between variables.

After the supervised learning methods, unsupervised learning approaches was applied to cluster variables across countries and the methods proposed to validate the results were K-means clustering, Hierarchy clustering, and Principal Component Analysis (PCA). Correlation calculations was performed among parameters used in the modeling step. An optimization method called Ordinary Least Square (OLS) for the epidemiology, socio-economic, geo-climatic and demographic determinants of COVID-19. Eventually, the deep learning methods used were Neural Network (NN) and Multi-Layer Perceptron (MLP) regressor, which is a class of feedforward Artificial Neural Network (ANN).

4.4.1 Socio-economic, geo-climatic and demographic variables

Socio-economic, geo-climatic and demographic variables constitute a strong determinant of the spread of the pandemic. These variables were extracted from [96, 160, 214, 217, 227, 228, 229, 230, 231], while some were calculated like the socio-economic fracture index. Some of the observed variables used are immigration rate (IR), average life expectancy (LE), Tuberculosis incidence (TB), temperature, percentage of gross domestic product devoted to health expenditure (CHE/GDP), percentage of 10% lowest (LI) and 10% highest incomes (HI), government response stringency index (SI), sustainable development goal (SDG) index, human development index (HDI), environmental performance index (EPI), consumer confidence index (CCI), stringency index (SI), Theil index (TI), and Gini index (GI). Other variables used are in Table 4.6, precisely for Cameroon regions. The data was collated based on the available countries and most recent years in public databases [96, 160, 214, 217, 227, 228, 229, 230, 231]. The calculated socio-economic variables are as follows:

- Social fracture (SF) index is the ratio between the 10% highest income and the 10% lowest income. In brief, it is expressed by the equation below:

$$\mathrm{SF} = \frac{10\%\mathrm{HI}}{10\%\mathrm{LI}}$$

- Demo-economic (DI) index is the ratio between the percentage of GDP devoted to health expenditure and social fracture index. It is expressed by the equation below:

$$\mathrm{DI} = \frac{\mathrm{CHE/GDP}}{\mathrm{SF}}$$

The precise value of all variables are in Table 4.8, Table 4.9, Table 4.10, Table A.2 in the Appendix A, Table B.1 and Table B.2 in the Appendix B.

4.4.2 Epidemiology variables

The epidemiologic variables are: first wave maximum R_o, second wave maximum R_o, first wave deterministic R_o, second wave deterministic R_o, maximal reproduction number $R_{\max}$,

transmission rate β and opposite of the initial autocorrelation slope averaged on 6 days for both first and second wave of the daily new cases for developed and developing countries. Other epidemiologic variables values are defined in Table 4.6.
The epidemiologic variables were recorded during the exponential phase of the first and second wave of the pandemic. Daily new cases observed during the first 100 days were used to calculate the exponential slope for the first and second wave. The opposite of the initial autocorrelation slope was averaged on 6 days for the first and second wave. The maximum R_o was collated from Table B.1 in the Appendix B while observing this value during the first and second waves of countries considered. Also collated from Table A.2 in the Appendix A is the deterministic R_o for the first and second wave of the pandemic taking 6 days as length of contagiousness period.
In this present study, the results were validated by performing cross-validation and also training 80% of the data and training 30%.

4.4.3 The relationship between Theil index and Gini index: mathematical and statistical approach

Firstly, the relationship between Theil index and Gini index is proved mathematically. The Gini index is defined as follows [96]:

$$\mathrm{GI} = 1 - \sum_{k=1,\ldots,n} (x_k - x_k - 1)(y_k - y_k - 1) = 1 - E(\Delta) = E(1 - \Delta),$$

where x_k(resp. y_k) denotes the kth cumulative part of the population (respectively income). On choosing the population increments, $d_k = x_k - x_{k-1}$ are equal to $1/n$, and if $E(\Delta)$ represents the expectation of the increment, $\Delta_k = y_k - y_{k-1}$ for the distribution d_k. Then, the Theil index applied to the percentage y_k of the total income relative to a percentage x_k of the total population ([96]) is defined by the following equation:

$$\mathrm{TI} = -\sum_{k=1,\ldots,n} (y_k - y_k - 1)\log(y_k - y_k - 1) = -\sum_{k=1,\ldots,n} \Delta k \log(\Delta k).$$

If the first increment of $y, \Delta_1 = y_1 \leq 1$, is close to 1 [which corresponds to a square-shaped Lorenz curve, i.e., close to a left right triangle-shaped income vs. population curve (in red on Figure 4.28), or to a high Gini index close to 1], then: $-\log(\Delta_1) \sim 1 - \Delta_1$ and then:

$$TI \sim -\log(\Delta_1) \sim 1 - \Delta_1 GI,$$

the equality being available only if the Lorenz curve presents a perfect left right triangle shape.

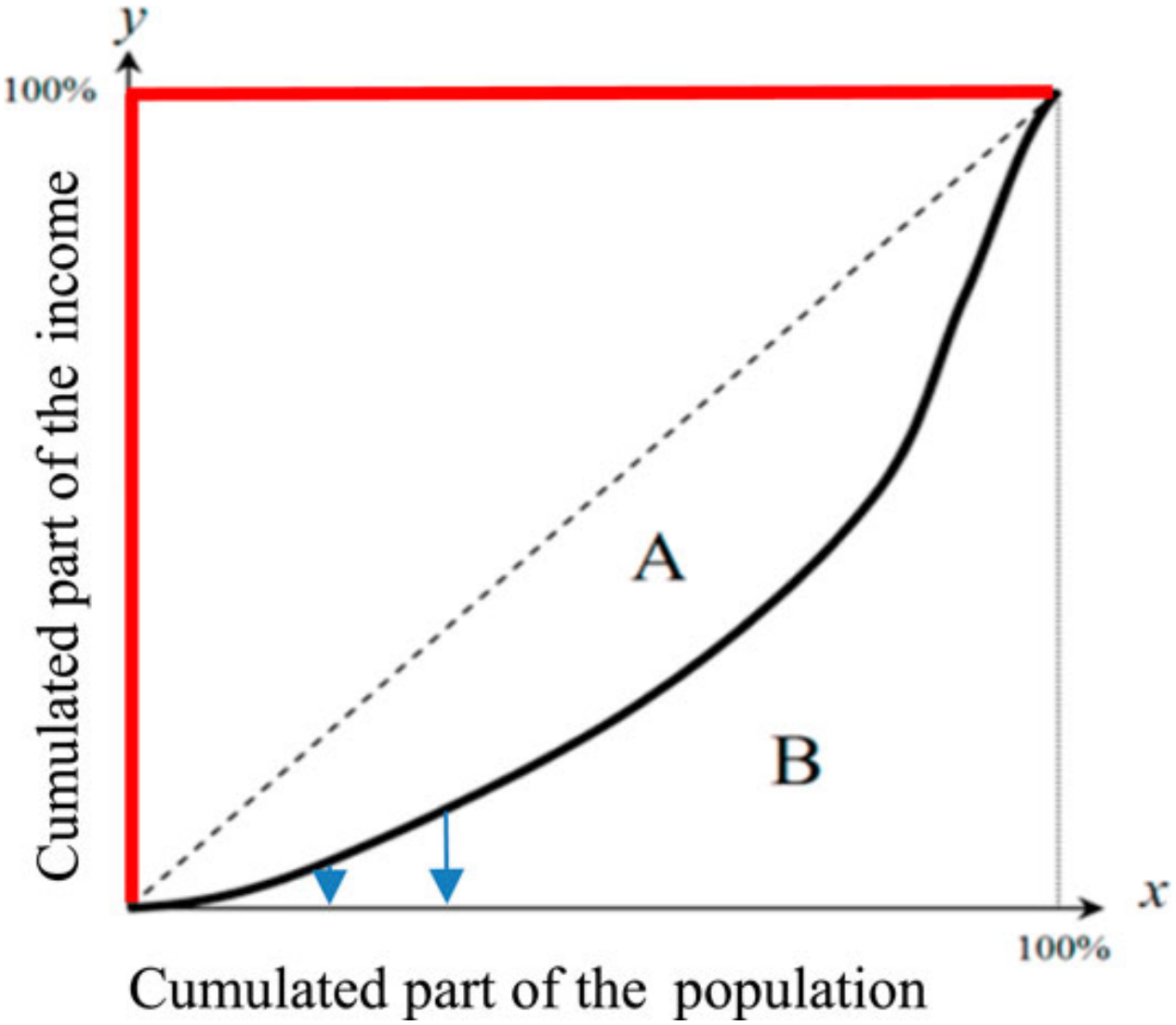


Figure 4.28: Lorenz curve showing the cumulated part of income vs. cumulated part of a population having this cumulated income. The curve in red represents left right triangle-shaped Lorenz curve.

Lastly, the statistical approach is in two folds: (1) to visualize the weights of the neural network and the bias between Theil and Gini index, and as it can be seen in Figure 4.29H, the weights are good with low bias and (2) to use linear and polynomial regression model of order 2 in order to see the relationship between Theil index and Gini index.

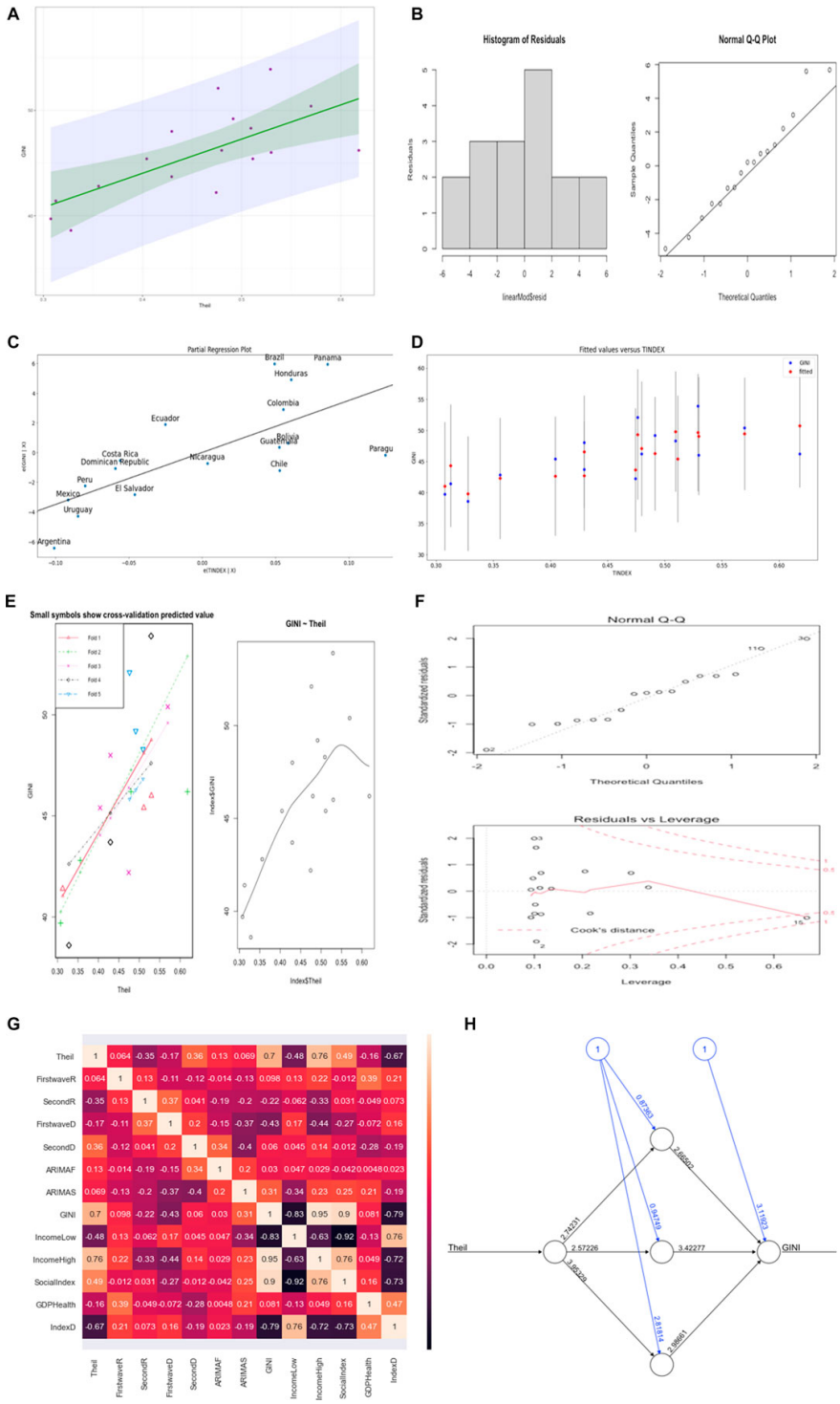


Figure 4.29: (A) Linear regression line (in green) with the confidence interval (in gray). (B) Residual plot for the linear regression. (C) Partial regression plot. (D) Fit plot. (E) On left-hand side is the cross-validation plot for the linear regression and on right-hand side is the polynomial regression plot. (F) Residual plots for polynomial regression. (G) Heat map for the correlations between all variables. (H) Neural network visualization.

For the linear regression as shown in Figure 4.29A, the intercept is 31.03, p-value is 0.0181, R-squared is 0.4881, residual standard error is 3.116, and all coefficients are significant with p-value < 0.05 for both the train and test data for linear and polynomial regression. The median

of the residual plot in Figure 4.29B and Figure 4.29F are 0.2111 and 0.2566, respectively, for both linear and polynomial regression, which are low values. The normality of the residual was tested using Jarque-Bera and Durbin-Watson tests, which gave a high p-value, and the null hypothesis was rejected, that is, the skewness and kurtosis of the residuals are statistically equal to zero. In order to know the performance of the linear regression model, 80% of the data was trained and 20% was tested and also did cross-validation to be sure of the accuracy. The predicted and the observed values are very close to the results presented for the regression models used. For the linear model, the cross-validation result is in Figure 4.29E whose average mean square error for the five portion folds is 11.72794. The correlation between the tested and the predicted values has high accuracy (R-squared = 0.97). The test set p-value is 0.02 with a residual standard error of 3.528. For polynomial regression of order 2, the train set has the following results: R-squared = 0.6, p-value = 0.002, and residual standard error = 2.935. The test set has the following results: R-squared = 0.99, p-value = 0.008, and residual standard error = 0.5639.

4.4.4 Correlations between epidemiologic, socio-economic and demographic parameters

In this section, two sets of correlation is done: (i) correlation between Theil and Gini indices with all epidemiologic, demographic, and socio-economic variables for Latino-American countries and (ii) correlation between epidemiologic, demographic, and socio-economic variables for Cameroon regions.
In Figure 4.29G, Theil and Gini indices are highly positively correlated with coefficient 0.7. By observing the epidemiologic variables (numbers of weekly reported new infectious cases, cured and deceased patients) during the 19 first weeks of 2021 (cf. in Figure 3.19, data from [119]). The parameters list is given in Table 4.6 and their values are extracted from Table 3.4 and from [126, 227]. The correlations between the parameters of Table 3.2 considered as variables on the 10 regions of Cameroon can be analyzed thanks to Figure 4.30 and Figure 4.31, which show that the epidemiologic parameters (from parameter Dose 1 to parameter Recovered) are significantly anti-correlated (pink color) with many socio-economic and demographic parameters, namely the three first ones (Subnational Human Development Index, International Wealth Index and Gross National Income per capita) and to those linked to the house holding, showing the influence of a good socio-economic state on the resistance against the virus. By exploiting the data corresponding to the parameters of Table 4.6, it is possible to analyze the correlations existing between three families of parameters observed in the different regions of Cameroon during the second wave of the COVID-19 outbreak.
In Figure 4.30, it can be observed that the socio-economic factors are highly positively correlated with each other, but negatively correlated with demographic parameters and not correlated with epidemiologic parameters, except those corresponding to vaccination and death which are anticorrelated with the signs of wealth. This can be explained by a higher socio-cultural level, which pushes individuals to monitor their health more and to get vaccinated. In Figure 4.31, it can be deduce that a greater number of demographic parameters are anti-correlated with socio-economic factors. These are in particular infant mortality, mean age, fertility and poverty indicators, which have positive or zero correlation coefficients between them and very negative with wealth indicators, such as GNI or possession of material goods or land (television and house), because in general a rise in the standard of living causes fertility and infant mortality decrease.

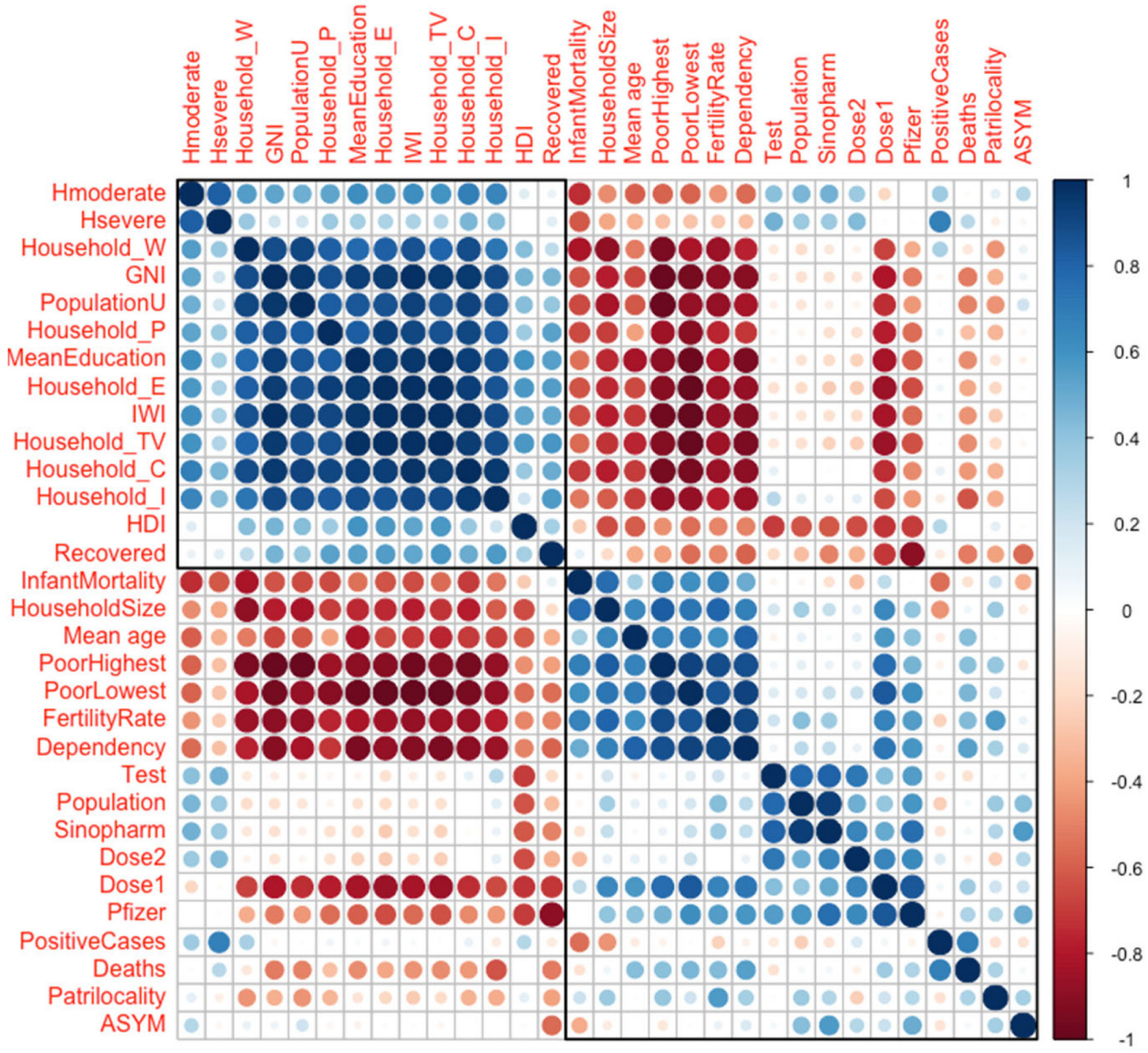


Figure 4.30: Correlations between various epidemiologic, socio-economic and demographic parameters observed in the different regions of Cameroon during the second wave. The false color scale indicates the intensity of the correlation between -1 and 1.

In Figure 4.34 a high correlation between Sud-Ouest and Littoral is observed with the score 0.918 and the least correlations are between Center and Adamaoua with the score 0.242, between Est and Center with the score 0.282 and between Nord and Center with the score 0.253. In Figure 4.32 there is high correlations between Ouest and Center with score 0.9308 and also between Littoral and Nord-Ouest with score 0.936967. The least correlation is between Extreme-Nord and Sud with score 0.04. In Figure 4.33 there is three anticorrelations: (i) between Nord and Adamaoua with score -0.00967, (ii) between Nord and Sud-Ouest with score -0.18, and (iii) between Sud-Ouest and Ouest with score -0.09988. The highest correlation is Nord-Ouest and Center with score 0.845.

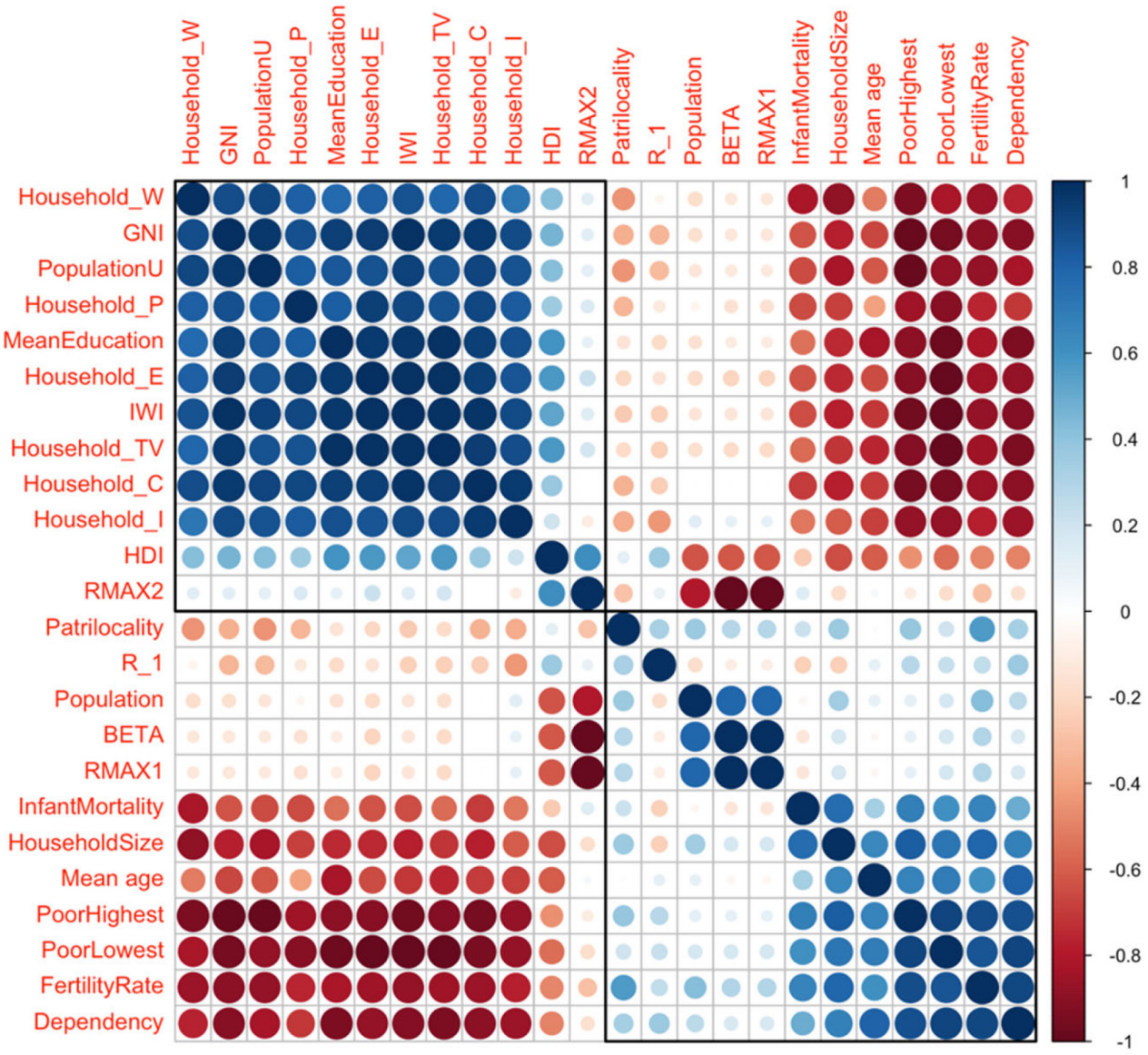


Figure 4.31: Correlations between various epidemiologic, socio-economic and demographic parameters observed in the different regions of Cameroon during the second wave. The false color scale indicates the intensity of the correlation between -1 and 1.

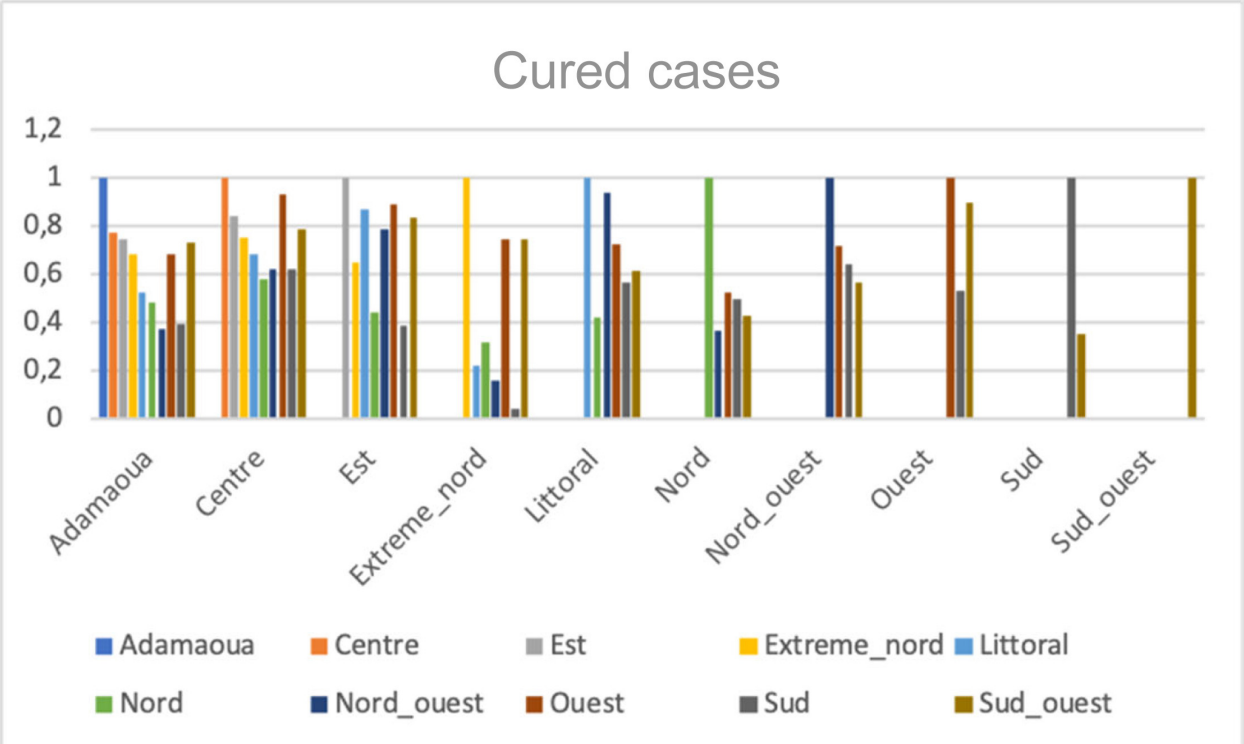


Figure 4.32: Correlations of the cured cases numbers between the 10 regions of Cameroon.

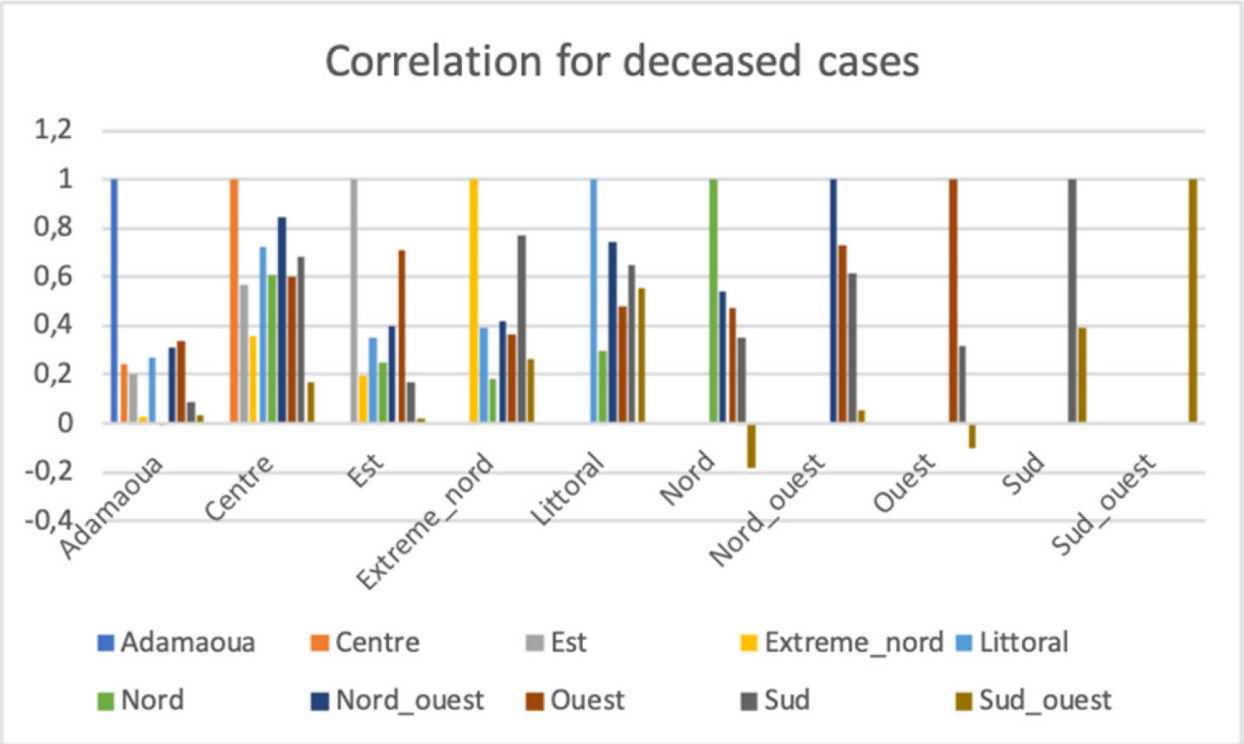


Figure 4.33: Correlations of the deceased cases numbers between the 10 regions of Cameroon.

Table 4.6: List of the parameters considered for the different regions of Cameroon.

Socio-economic parameters	
HDI	Subnational Human Development Index
IWI	International Wealth Index
GNI	Gross National Income per capita
PoorHighest	Percentage poor households (IWI value under 70)
PoorLowest	Percentage poor households (IWI value under 35)
MeanEducation	Mean years education population aged 20+
Household-TV	Percentage households with a TV
Household-C	Percentage households with a computer
Household-W	Percentage households with piped water
Household-E	Percentage households with electricity
Household-P	Percentage households with a phone
Household-I	Percentage households with internet access
HouseholdSize	Average household size
Patrilocality	Patrilocality index (positive values patrilocal)
Dependency	Dependency ratio
Demographic parameters	
FertilityRate	Total fertility rate
InfantMortality	Infant mortality rate
Population	Total area population in millions
PopulationU	Urban population size
Epidemiologic parameters	
Dose1	Number of vaccinated people that have received first dose
Dose2	Number of vaccinated people having received second dose
Pfizer	Number of vaccinated people with Pfizer vaccine
Sinopharm	Number of vaccinated people with Sinopharm vaccine
Test	Number of tested people
PositiveCases	Number of people that tested positive
ASYM	Patients that are asymptomatic and mild symptoms
Hmoderate	Patients with moderate symptoms
Hsevere	Patients with severe symptoms
Deaths	Number of deaths recorded
Recovered	Number of recovery recorded
R_1	Daily reproduction number of the first day of infection
$R_{\max_1}$	R maximum first wave
$R_{\max_2}$	R maximum second wave
BETA	Transmission rate

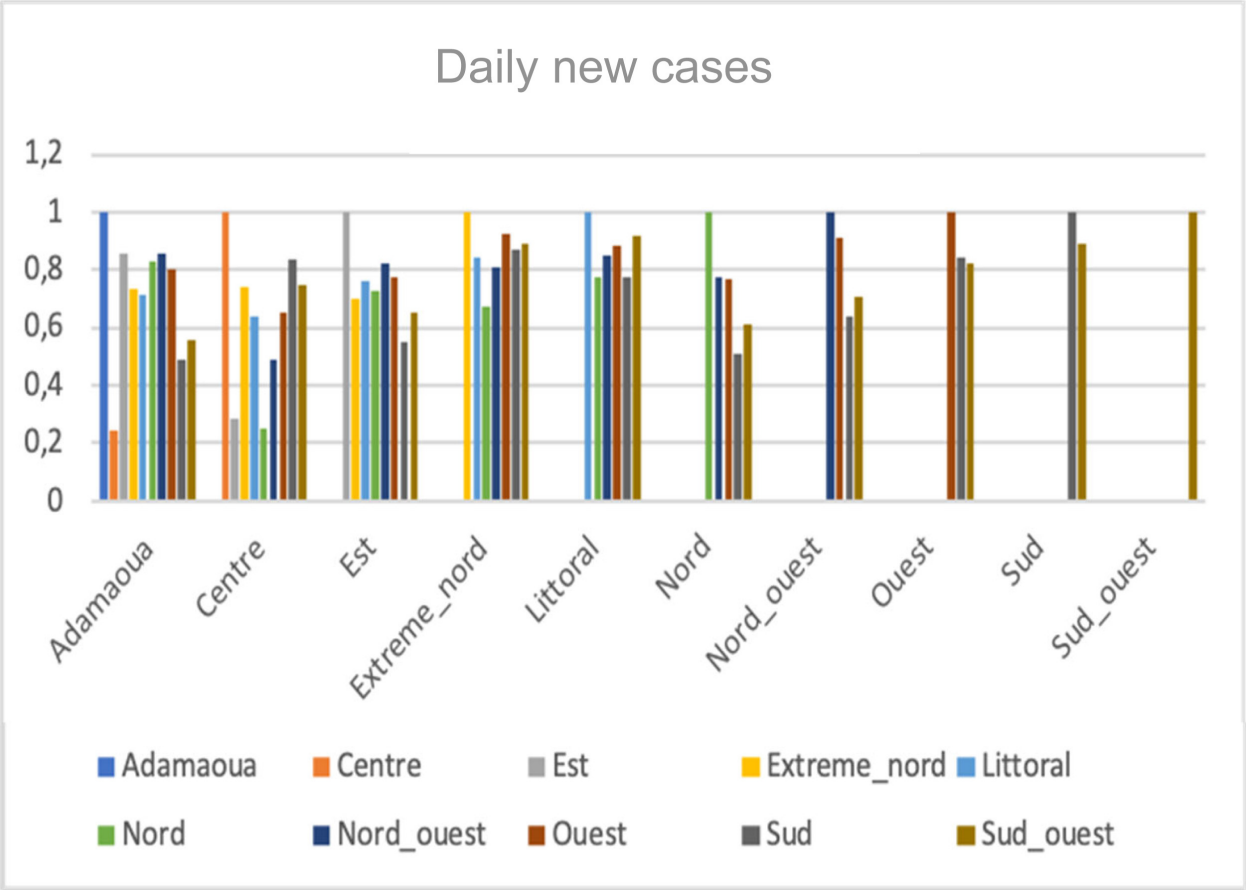


Figure 4.34: Correlations of the new cases numbers between the 10 regions of Cameroon.

4.4.5 Multivariate analysis, clustering and prediction for Latino-American countries

The multivariate analysis using ordinary least square method, clustering analysis using PCA, K-means and Hierarchy clustering and the prediction of the Gini index using MLP regressor which is a deep learning method, linear, lasso and ridge regression for Latino-American countries is presented.

Figure 4.37 corresponds to the ordinary multivariate least square methods with R-squared = 0.674. Figure 4.37A shows Paraguay as outlier not fitting data, Figure 4.37B normalizes all countries and does not point any country in the plot. Using cross-validation method, the best parameter α for the modeling is shown in Figure 4.35C. For ridge regression, α=0.142 with a mean square error of 1.36 and α=0.368 for lasso regression with a mean square error = 5.10. For Figure 4.35E, training score = 1.000 and test score = 0.641; for Figure 4.35F, training score = 0.992 and test score = 0.497; for Figure 4.36A, training score = 0.99 and test score = 0.406; and for Figure 4.36B, training score = 0.984 and test score = -0.077. It is evident from these results that linear regression best predicts Gini index with the highest test score, and predicted values are very close to each other as presented in Table 4.7. Also, the same pattern of prediction is observed in Figure 4.35E–F and Figure 4.36A-B showing that all methods used in this section have the same predictive behavior.

Figure 4.38C, the first two clusters have 14 countries and the third has three countries, which are Uruguay and El Salvador on the same hierarchy while Argentina is on another hierarchy.

Only the first cluster in the dendrogram is presented. In Figure 4.38F, the Gini index has the highest positive correlation of 0.44 with the principal component PC 1 and Theil index has only the value 0.34 with PC 1. The main variable causing the separation into three classes is the Gini index.

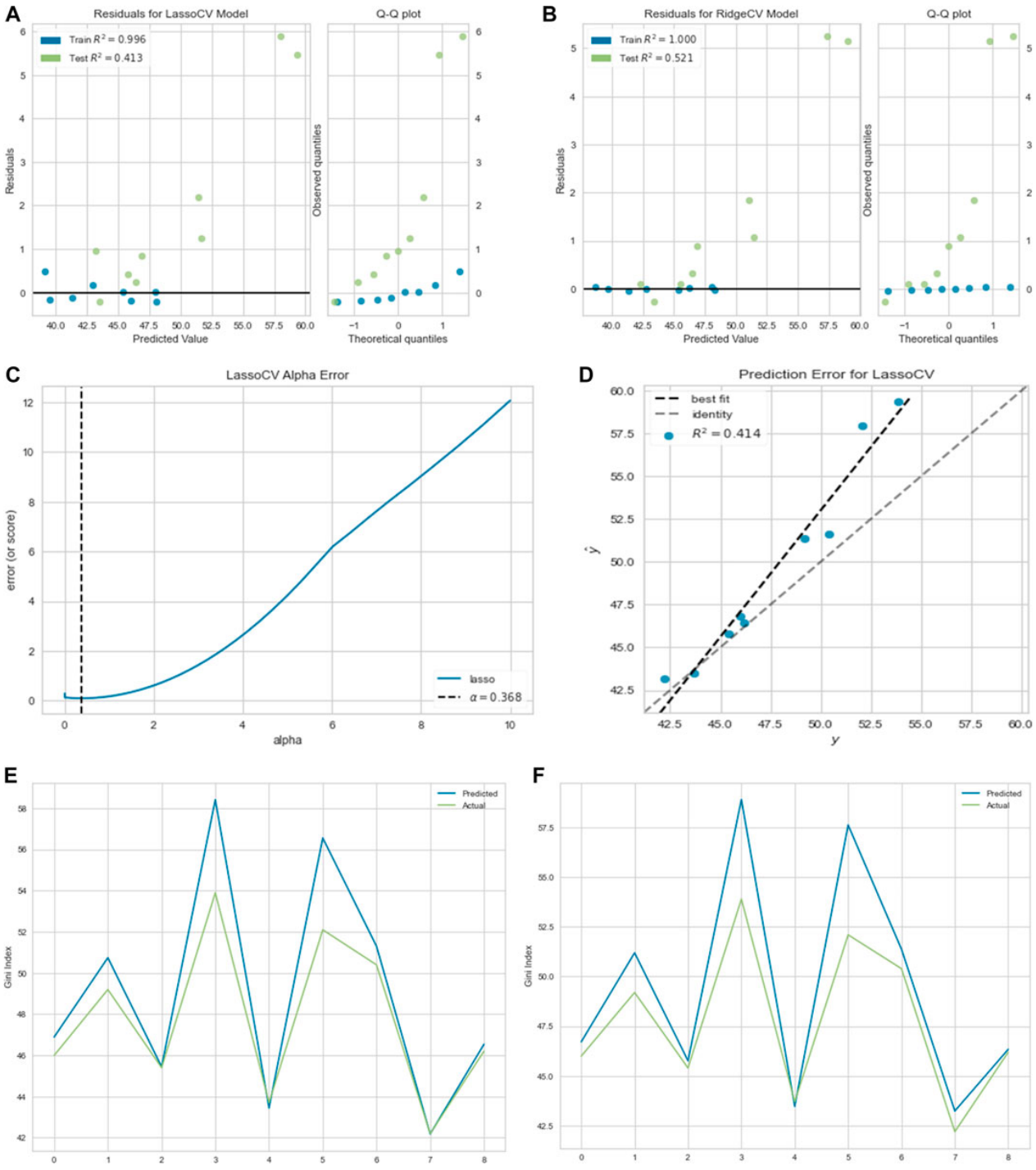


Figure 4.35: (A) Residual plot for lasso regression. (B) Residual plot for ridge regression. (C) Lasso regression cross-validation error. (D) Prediction error for the lasso regression. (E) Linear regression prediction plot. (F) Lasso regression and (G) ridge regression prediction plots. (H) MLP regression prediction plot.

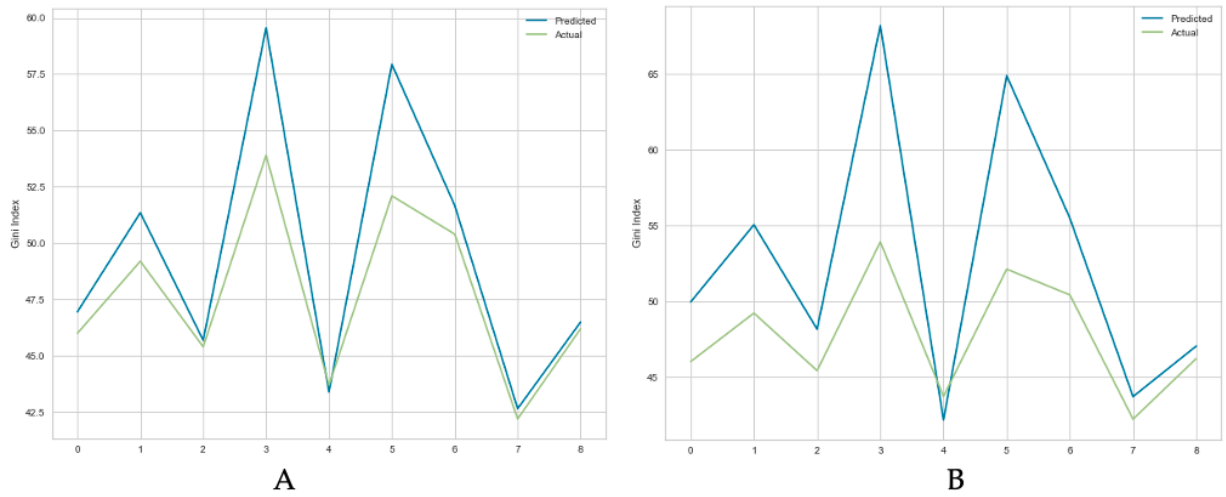


Figure 4.36: (A) Ridge regression prediction plots. and (B) MLP regression prediction plot.

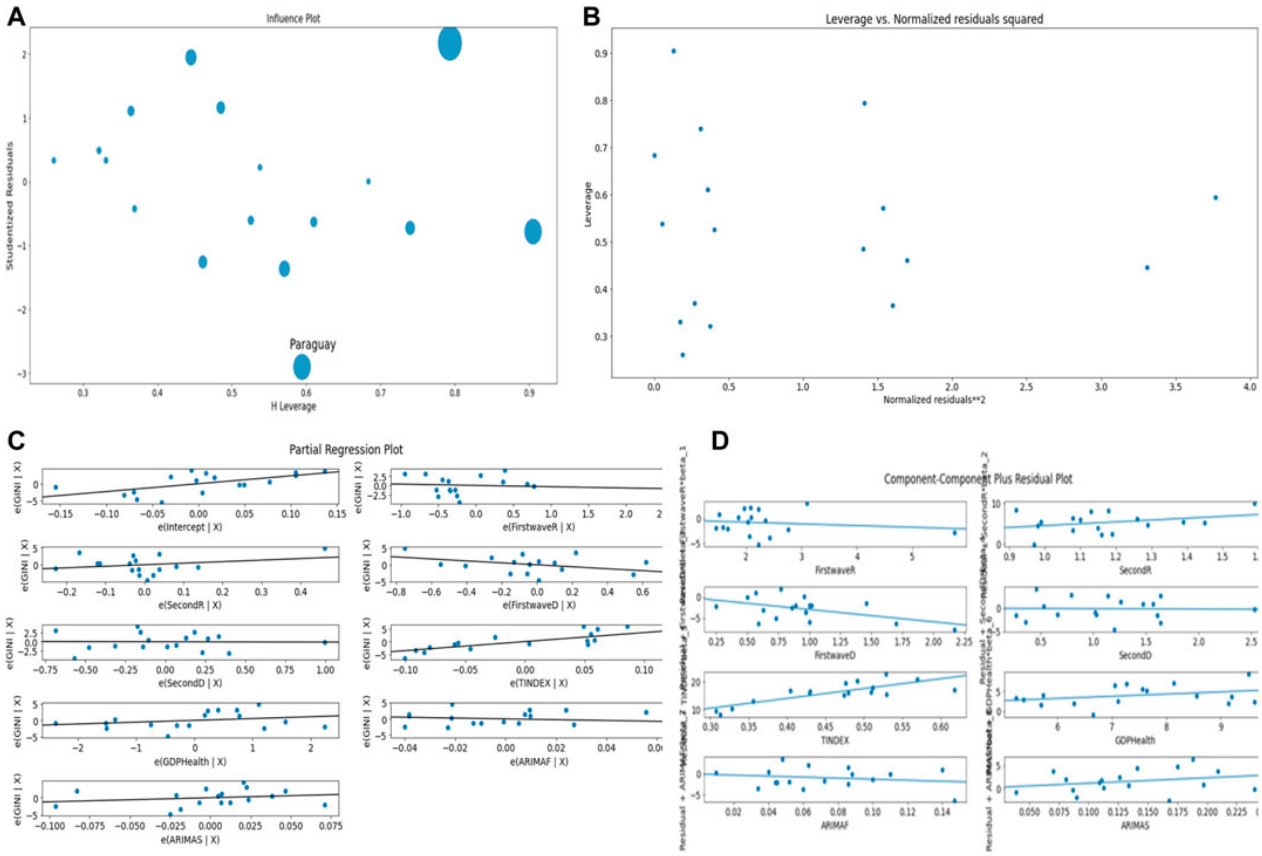


Figure 4.37: (A) Influence plot. (B) Leverage vs. Normalized residuals squared plot. (C) Partial regression plot. (D) Component–Component plus residual plot.

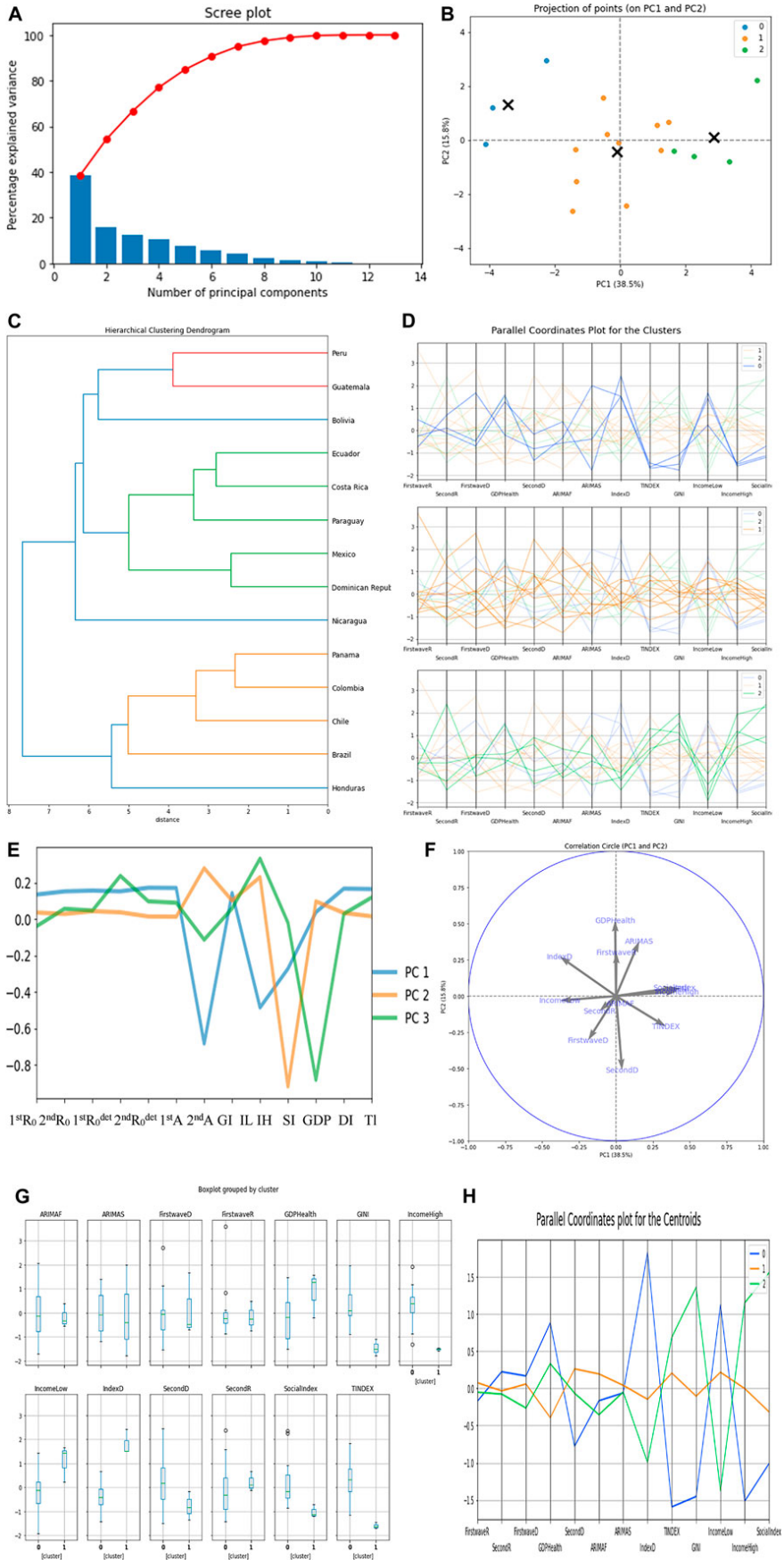


Figure 4.38: (A) Scree plot. (B) Plot for projection of points for PC1 and PC2. (C) Hierarchy clustering dendrogram. (D) Parallel coordinates plot for the clusters. (E,F) PC's visualization. (G) Box plot for the clusters. (H) Parallel coordinates plot for the centroids.

Table 4.7: Comparison of different regression models for the prediction.

S/N	Country Name	Gini Index	Linear Regression	Lasso Regression	Ridge Regression	MLP Regressor
1	PARAGUAY	46.2	46.5	46.3	46.5	46.0
2	PANAMA	49.2	50.7	51.2	51.3	51.5
3	BRAZIL	53.9	58.4	58.9	59.6	61.7
4	BOLIVIA	42.2	42.2	43.2	42.7	42.0
5	HONDURAS	52.1	56.6	57.6	-	59.6
6	DOMINICAN	43.7	43.4	43.5	43.4	43.8
7	CHILE	46.0	46.9	-	47.0	48.3
8	MEXICO	45.4	-	45.8	45.7	-
9	COLUMBIA	50.4	51.3	51.4	51.7	52.8

Table 4.8: Latino-American countries data.

Country Name	Gini index	10% LI	10% HI	SF index	Demographic index	Theil index
Argentina	41.4	1.8	29.9	16.61	0.58	0.312583
Bolivia	42.2	1.5	30.4	20.27	0.31	0.474293
Brazil	53.9	1.0	42.5	42.50	0.22	0.528996
Chile	46.0	1.8	37.1	20.61	0.44	0.529734
Colombia	50.4	1.4	39.7	28.36	0.27	0.569899
Costa Rica	48.0	1.5	36.3	24.20	0.31	0.429539
Dominican	43.7	2.3	35.2	15.30	0.37	0.429447
Ecuador	45.4	1.6	34.4	21.50	0.38	0.404237
El Salvador	38.6	2.4	29.4	12.25	0.58	0.327834
Guatemala	48.3	1.7	38.0	22.35	0.26	0.509287
Honduras	52.1	0.9	39.1	43.44	0.16	0.476235
Mexico	45.4	2.0	36.4	18.20	0.30	0.511219
Nicaragua	46.2	2.0	37.2	18.6	0.46	0.479821
Panama	49.2	1.2	37.1	30.92	0.24	0.491384
Paraguay	46.2	1.7	35.9	21.12	0.31	0.618109
Peru	42.8	1.8	32.1	17.83	0.29	0.355739
Uruguay	39.7	2.3	29.7	12.91	0.71	0.307564

4.4.6 Multivariate analysis, clustering and prediction for developed and developing countries

The multivariate analysis using ordinary least square method, clustering analysis using PCA, K-means and Hierarchy clustering and the prediction of the percentage GDP devoted to health expenditure using MLP regressor which is a deep learning method, linear, lasso and ridge regression for developed and developing countries is presented.

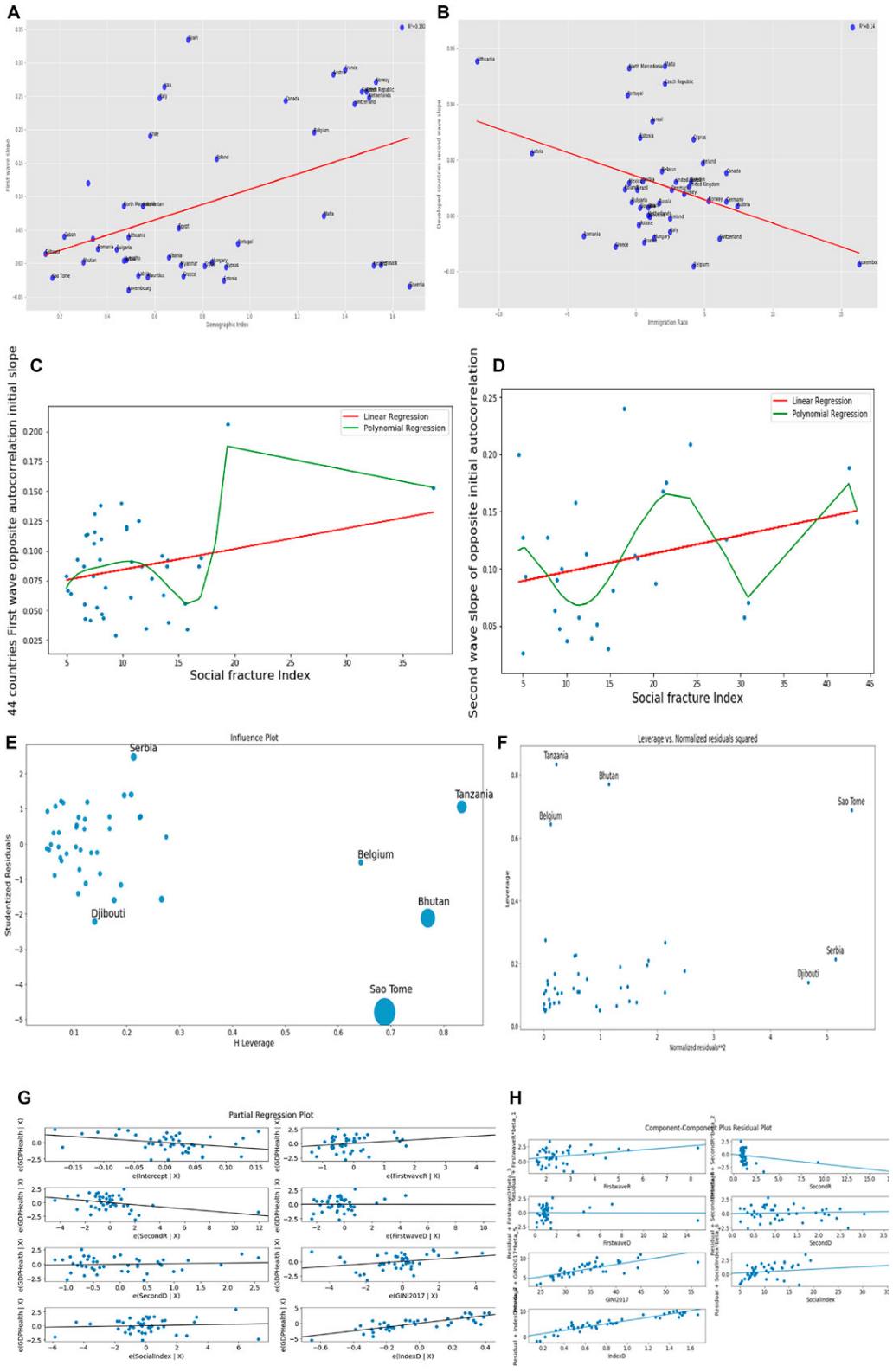


Figure 4.39: Linear regression plots for (A) the first wave slope vs. the demo-economic index for the developed and developing countries, (B) second wave slope vs. immigration rate for developed countries, (C) the opposite of the initial auto-correlation slope for first wave vs. social fracture index, (D) opposite of initial autocorrelation slope for second wave vs. social fracture index. (E) Influence plot. (F) Leverage vs. normalized residuals squared plot. (G) Partial regression plot. (H) Component–Component plus residual plot.

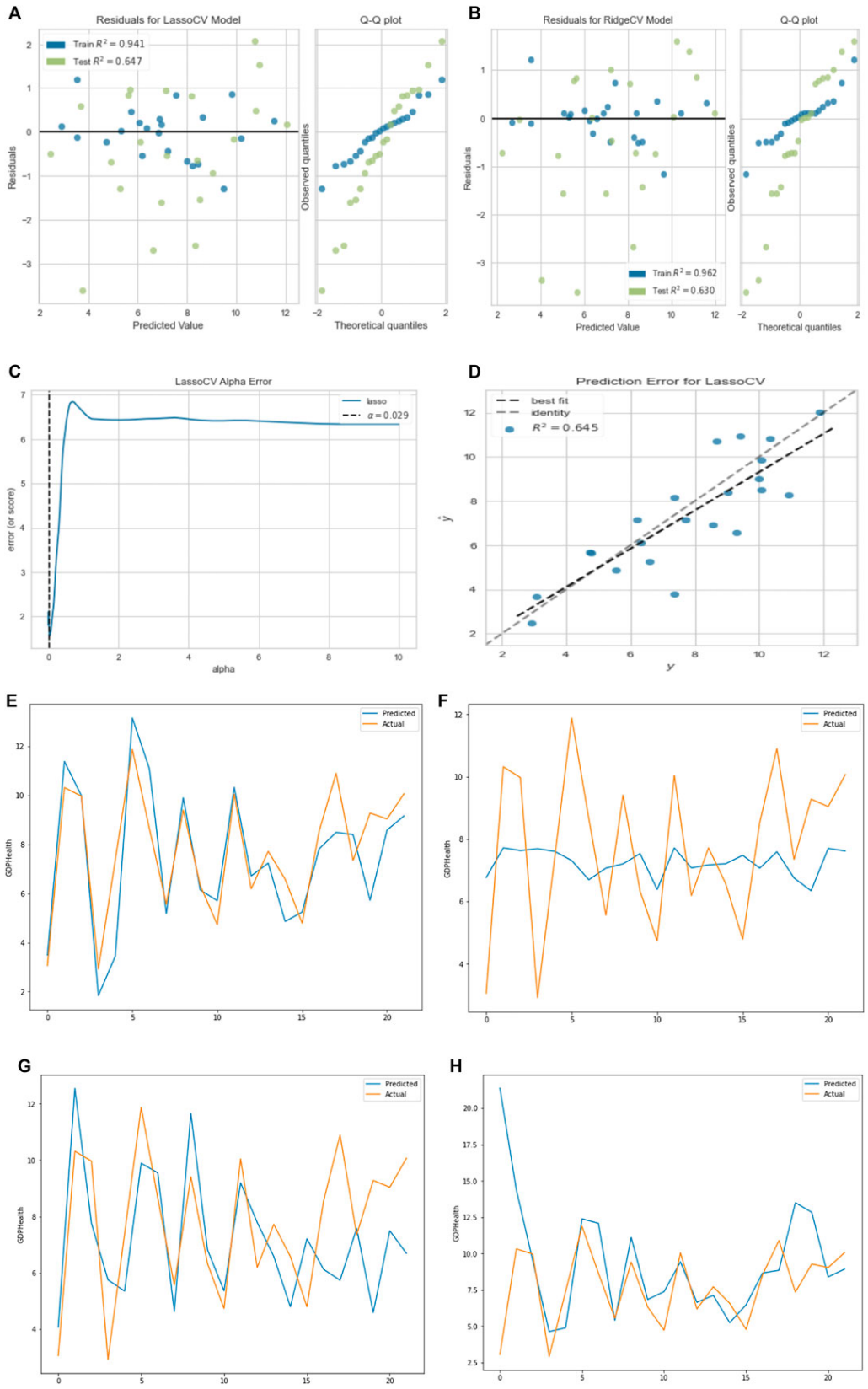


Figure 4.40: (A) Residual plot for lasso regression. (B) Residual plot for ridge regression. (C) Lasso regression cross-validation error. (D) Prediction error for lasso regression curve (R = 0.8). (E) Linear regression prediction plot. (F) Lasso regression prediction plot. (G) Ridge regression prediction plot. (H) MLP regression prediction plot.

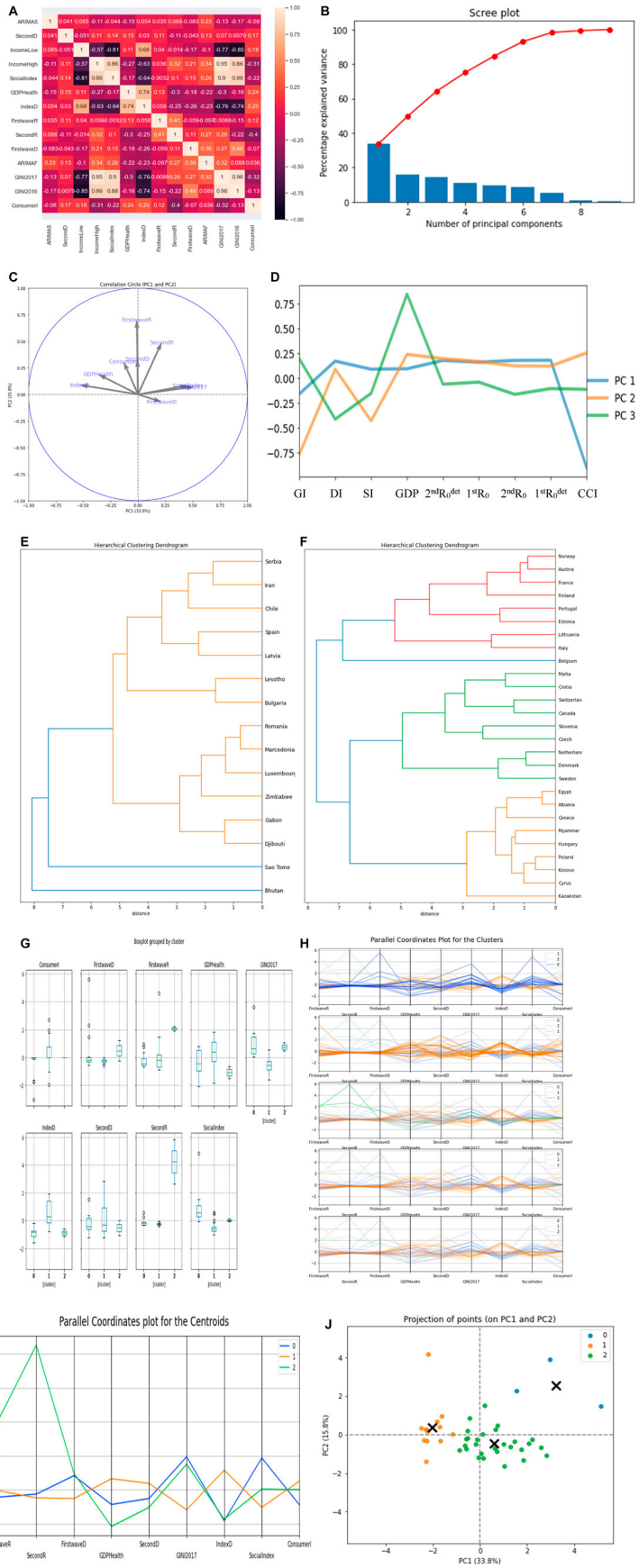


Figure 4.41: (A) Heatmap of the parameter's correlations. (B) Scree plot. (C,D) PCs visualization. (E,F) Hierarchy clustering dendrogram. (G) Boxplot of the clusters. (H) Parallel coordinates plot for the clusters. (I) Parallel coordinates plot for the centroids. (J) Projection points for PC1 and PC2.

Table 4.9: Developed and developing countries data.

Country Name	10% LI	10% HI	SF	DI	CCI	GI	Country Name	Immigration rate
ALBANIA	3.1	24.8	8.00	0.66	-	33.2	AUSTRIA	7.4
AUSTRIA	3.0	23.0	7.67	1.35	99.15	29.7	BELARUS	1.9
BELGIUM	3.3	21.9	8.11	1.27	100.75	27.4	BELGIUM	4.2
BHUTAN	2.7	27.9	10.33	0.30	-	37.4	BRAZIL	0.1
BULGARIA	1.9	31.9	16.79	0.44	-	40.4	BULGARIA	-0.3
CANADA	2.7	25.3	9.37	1.15	-	33.3	CANADA	6.6
CHILE	2.3	36.3	15.78	0.58	96.55	44.4	CHINA	-0.2
CROATIA	2.7	22.8	8.44	0.81	-	30.4	CYPRUS	4.2
CYPRUS	3.4	25.5	7.50	0.90	-	31.4	CZECH	2.1
CZECH	4.2	21.5	5.12	1.49	98.97	24.9	DENMARK	2.6
DENMARK	3.7	24.0	6.49	1.55	100.50	28.7	ESTONIA	0.3
DJIBOUTI	1.9	32.3	17.00	0.14	-	41.6	FINLAND	2.5
EGYPT	3.8	26.9	7.08	0.70	-	31.5	FRANCE	0.6
ESTONIA	3.0	22.5	7.50	0.89	97.59	30.4	GERMANY	6.6
FINLAND	3.8	22.6	5.95	1.52	100.27	27.4	GREECE	-1.5
FRANCE	3.2	25.8	8.06	1.40	98.12	31.6	HUNGARY	1.3
GABON	2.2	27.7	12.59	0.22	-	38.0	IRELAND	4.9
GREECE	2.4	25.9	10.79	0.72	98.11	34.4	ISRAEL	1.2
HUNGARY	3.0	23.9	7.97	0.84	98.88	30.6	ITALY	2.5
IRAN	2.3	31.3	13.61	0.64	-	40.8	KOSOVO	-
ITALY	1.9	26.7	14.05	0.62	100.25	35.9	LATVIA	-7.6
KAZAKHSTAN	4.3	23.0	5.35	0.55	-	27.5	LITHUANIA	-11.6
KOSOVO	3.8	24.6	6.47	-	-	29.0	LUXEMBOURG	16.3
LATVIA	2.3	26.9	11.70	0.53	95.05	35.6	MALTA	2.1
LESOTHO	1.7	32.9	19.35	0.48	-	44.9	MEXICO	-0.5
LITHUANIA	2.1	28.4	13.52	0.49	100.98	37.3	NETHERLANDS	0.9
LUXEMBOURG	2.4	25.8	10.75	0.49	98.52	34.9	MACEDONIA	-0.5
MALTA	3.4	23.3	6.85	1.31	-	29.2	NORWAY	5.3
MAURITIUS	2.9	29.9	10.31	0.57	-	36.8	POLAND	-0.8
MYANMAR	3.8	25.5	6.71	0.71	-	30.7	PORTUGAL	-0.6
N.LANDS	3.5	23.3	6.66	1.50	99.93	28.5	ROMANIA	-3.8
NORTH.M	1.7	23.8	14.00	0.47	-	34.2	RUSSIA	1.7
NORWAY	3.3	21.6	6.55	1.53	-	27.0	SERBIA	0.5
POLAND	3.2	23.5	7.34	0.86	98.70	29.7	SLOVAK	0.3
PORTUGAL	2.7	26.7	9.89	0.95	97.41	33.8	SLOVENIA	1.0
ROMANIA	1.6	24.9	15.56	0.36	-	36.0	SPAIN	0.9
SERBIA	1.4	25.6	18.29	0.47	-	36.2	SWEDEN	4.0
SAO TOME	1.3	49.1	37.77	0.17	-	56.3	SWITZERLAND	6.1
SLOVENIA	4.1	20.4	4.98	1.67	96.34	24.2	TURKEY	3.5
SPAIN	2.1	25.4	12.09	0.74	96.61	34.7	UKRAINE	0.2
SWEDEN	3.0	22.3	7.43	1.47	101.89	28.8	UK	3.9
SWITZERLAND	3.1	25.5	8.23	1.44	97.47	32.7	USA	2.9
TANZANIA	2.9	33.1	11.41	0.32	-	40.5	-	-
ZIMBABWE	2.5	34.8	13.92	0.34	-	44.3	-	-

- In Figure 4.39C and Figure 4.39D, the type of regression used is polynomial or order 6 while Figure 4.39A and Figure 4.39B is linear regression.
- The Figure 4.39 present regression analyses with the following parameters:
- Figure 4.39A:

  Linear regression: slope = 0.11463663009107196, intercept = -0.0037118697103040027, r-value = 0.4387157758684147, p-value = 0.0032517683682962654, standard error = 0.03667137141150123, R-squared = 0.192472 and RMSE = 0.1044724946057671.
- Figure 4.39B:
  - Linear regression: slope = -0.002547609589041096, intercept = 0.07888755616438356, r-value = -0.3272886357381478, p-value = 0.03672803730354382, standard error = 0.0011777868896598461, R-squared = 0.107118 and RMSE = 0.03065537183298402.
- Figure 4.39C:
  - Linear regression: slope = 0.0017309145398433248, intercept = 0.06695128460299407, r-value = 0.263675660748951, p-value = 0.08754941979369255, standard error = 0.0009889311191763849, R-squared = 0.069525 and RMSE = 0.0354860744891158,
  - Polynomial regression (order 6): R-squared = 0.3 and RMSE = 0.04060485094256808.
- Figure 4.39D:
  - Linear regression: slope = 0.0015999465132904799, intercept = 0.0810899892250729, r-value = 0.2861266574746827, p-value = 0.13239511534872409, standard error = 0.001031140187045727, R-squared = 0.081868 and RMSE = 0.05492215494302141,
  - Polynomial regression (order 6): RMSE = 0.07286590609946085 and R-squared = 0.35.
- Figure 4.39E–H correspond to the ordinary multivariate least square method with R-squared = 0.76. Figure 4.39E shows some developing countries as outliers, while Belgium is the only developed country, which does not fit the data. Cross-validation method helps to choose the best parameter α for the modeling as shown in Figure 4.40C. For ridge regression, $\alpha = 0.012$ with a mean square error = 2.32 and $\alpha = 0.029$ for the lasso regression with a mean square error = 2.21. For Figure 4.40E, training score = 0.983 and test score = 0.607; for Figure 4.40F, training score = 0.170 and test score = 0.021; for Figure 4.40G, training score = 0.854 and test score = 0.115; and for Figure 4.40H, training score = 0.980 and test score = -2.386.

  It is evident from the results that linear regression best predicts GDP percentage devoted to health expenditure with the highest test score, and all predicted values are very close. In Figure 4.41E and Figure 4.41F, the first cluster has 15 countries, the second cluster has 27 countries, while the last cluster has 2 countries, which are Tanzania and Mauritius. Only the first two cluster dendrograms are presented. With PC1, Gini index GI has the highest positive correlation of 0.52 with PC1 and demo-economic index DI

has the second highest negative correlation of -0.53, while with PC2, first wave maximum R_0 has the highest positive correlation of 0.70 (Figure 4.41C). The first cluster contains a majority of developing countries, and the second cluster contains a majority of developed countries, the main variable causing the separation into two classes being the Gini index in PC1.

Also, sextic regression (polynomial regression of order six) was used to studied the correlation between the value of the opposite of the slope at the origin of the autocorrelation function of the first wave and the economic and health index CHE/GDP, by studying a polynomial regression of degree 6 (Figure 4.42) whose results is as follows:

- Linear regression: slope = 0.01117, intercept = 0.0664, r-value = 0.47219, p-value = 0.0097, standard error = 0.004, RMSE = 0.0387,
- Polynomial regression (order 6): RMSE = 0.04399, R-squared = 0.4.

It shows an anti-correlation in the linear regression and a local maximum for countries with an average CHE/GDP ratio of around 7. Countries with a high CHE/GDP ratio (such as France and the United States) have a low value in opposite to this slope. The explanation for this phenomenon may come from the correlation reported in the introduction between the CHE/GDP and Gini indices, the poor classes having a longer duration of contagiousness due to a less important state of immunological defense and perhaps less compliance with mitigation measures.

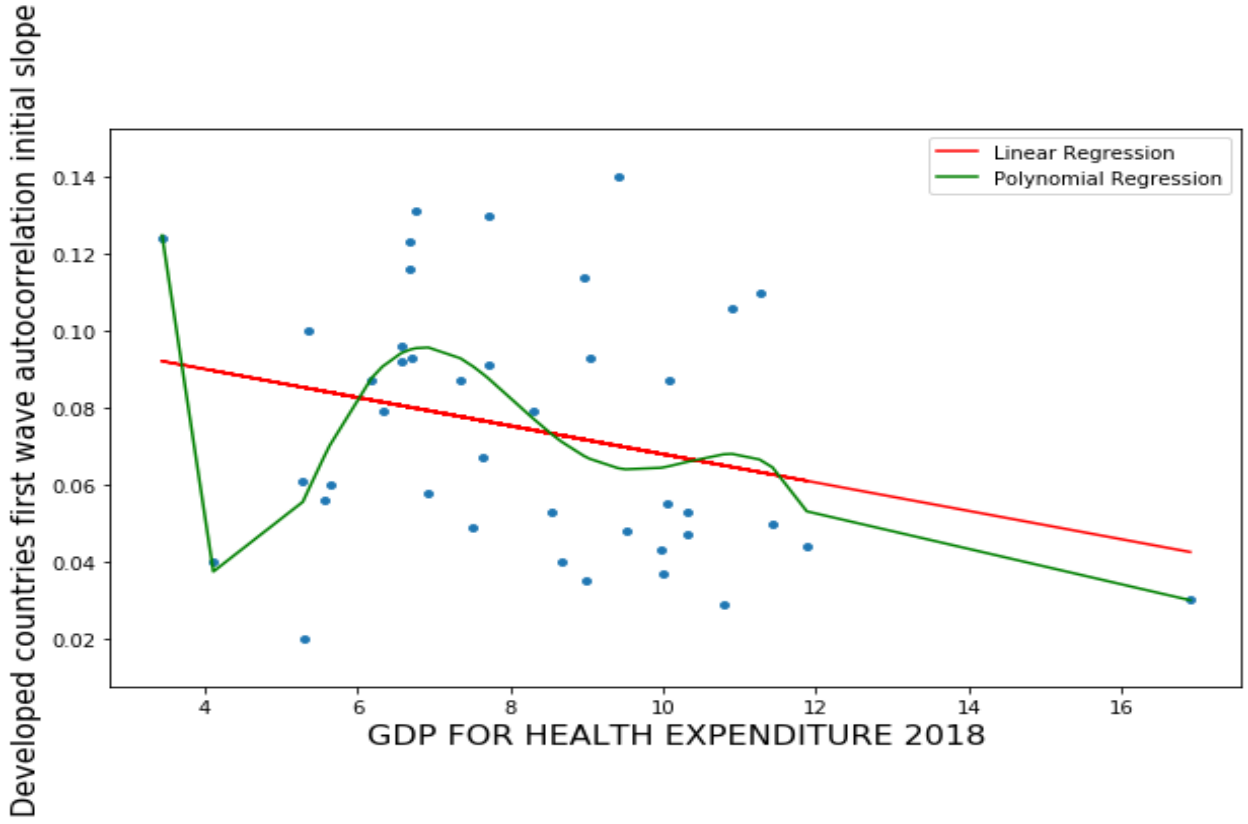


Figure 4.42: Linear (in red) and sextic (in green) regression plots of first wave opposite of initial autocorrelation slope vs. CHE/GDP.

- The correlation between the first wave exponential regression slope and the CHE/GDP index for developed and developing countries is significantly positive ($R = 0.57$) in Figure 4.44 with the result as follows:

  slope = 0.026632, intercept = -0.1052912, r-value = 0.5661, p-value = 7.60×10^{-5}, standard error = 0.00605655, R-squared = 0.320470, RMSE = 0.095836.

- In Figure 4.43, the regression analysis result is as follow:

  (**a**) slope = 0.0320468, intercept = -0.16158, r-value=0.6481, p-value = $4.62\,13 \times 10^{-6}$, standard error = 0.00603, R-squared = 0.42, RMSE = 0.09359760581,

  and (**b**) slope = -0.0010489, intercept = 0.01994, r-value = 0.1340845, p-value = 0.4094462, standard error = 0.001258, R-squared = 0.018, RMSE for linear regression = 0.018583749.

- Figure 4.43(a) and Figure 4.44 show a positive correlation while Figure 4.43(b) show a negative correlation between the slope of the logarithmic regression curve of the new cases of COVID-19 as a function of time (a sign of rapid growth of the epidemic if it is high) and the economic index CHE/GDP. This is true when we observe the developed and developing countries (Figure 4.44) or the developed countries alone for which the positive correlation is higher, the correlation coefficient increased from 0.57 to 0.65 (Figure 4.43(a)), but this trend is reversed for the second wave (Figure 4.43(b)), where the correlation coefficient equal -0.57, this being possibly due the early implementation of mitigation measures in developed countries, reducing the exponential growth of new cases in the second wave.

4.4.7 Multivariate analysis, clustering and prediction for African countries

The multivariate analysis using ordinary least square method, clustering analysis using PCA, K-means and Hierarchy clustering and the prediction of the temperature (which is a variable seen to influence the spread of the pandemic in this continent due to low cases and deaths observed) using MLP regressor which is a deep learning method, linear, lasso and ridge regression for African countries is presented.

- Figure 4.47A–D correspond to the ordinary multivariate least square method with $R^2 = 0.60$.

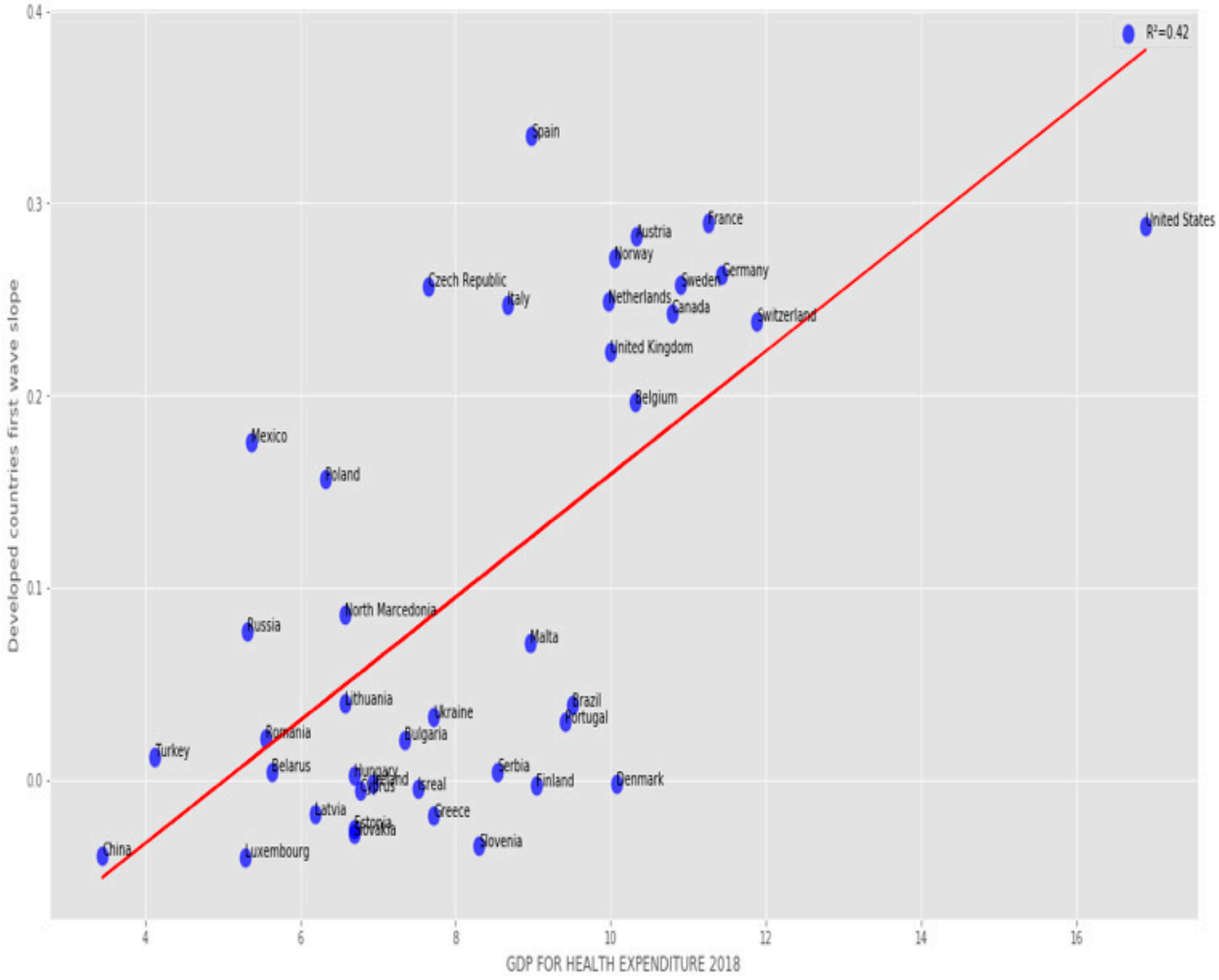
- Figure 4.47A shows Botswana and Tanzania as outliers not fitting the data. Cross-validation method helps to choose the best parameter α for the modeling as shown in Figure 4.45C.

- For ridge regression, $\alpha = 1.005$ with a mean square error of 19.13,

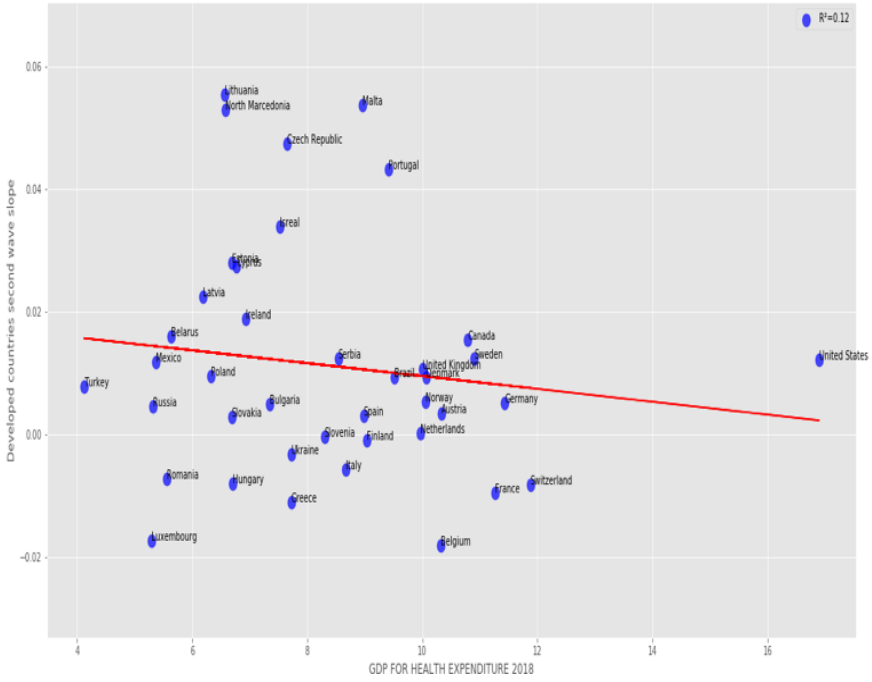
- For lasso regression, $\alpha = 6.018$ with a mean square error = 16.93. For Figure 4.45E, training score = 0.647 and test score = -2.228

- For Figure 4.45F, training score = 0.316 and test score = 0.154,

- For Figure 4.45G, training score = 0.573 and test score = -1.136,



(**a**)



(**b**)

Figure 4.43: Regression plots for developed countries of (**a**) first and (**b**) second wave exponential regression slope versus CHE/GDP. 182

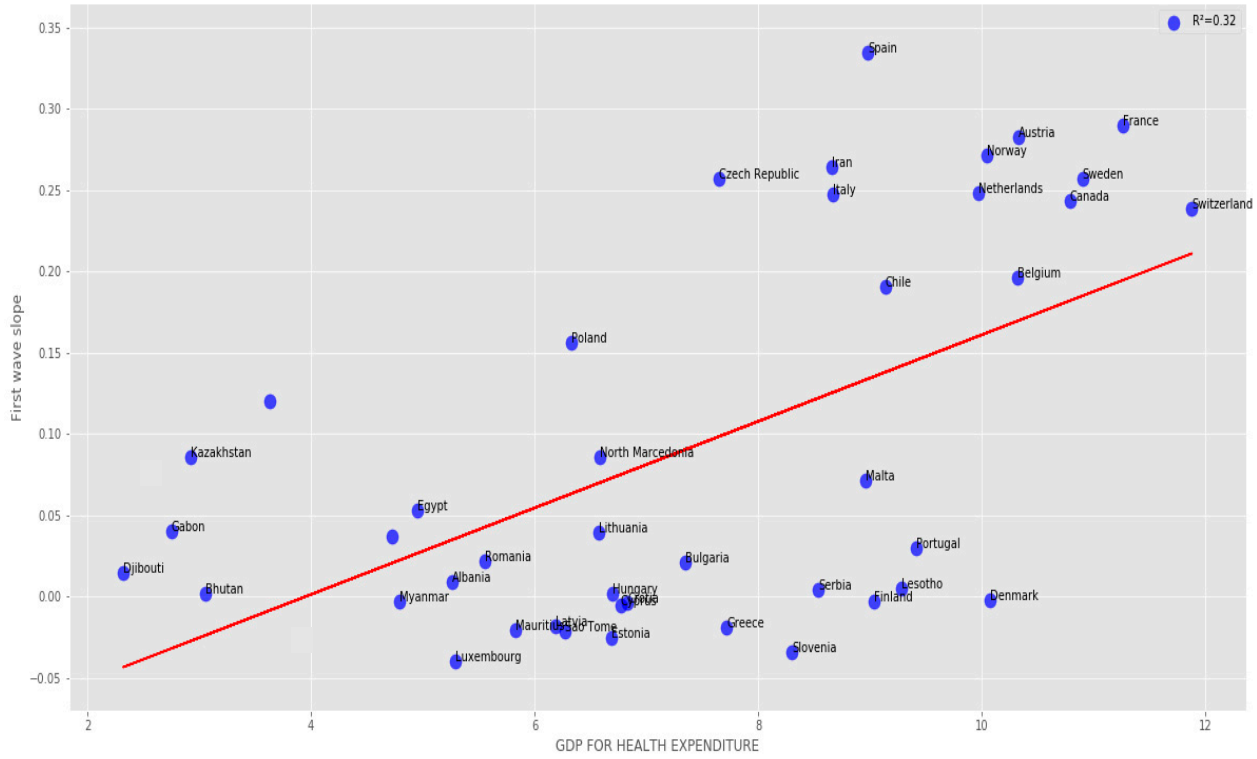


Figure 4.44: Regression plot of first wave exponential regression slope vs CHE/GDP for developed and developing countries.

- For Figure 4.45H, training score = -6.728 and test score = -4.714.

It is evident from these results that the lasso regression best predicts temperature with the highest test score, and predicted values of temperature for lasso and ridge regression are close. All the regression methods give about the same result with the maximum accuracy for the ridge regression. In Figure 4.46E and Figure 4.46F, the first cluster has 40 countries, the second cluster has 13 countries, while the last cluster has only 1 country, which is Botswana. Only two cluster dendrograms with many countries are presented. In Figure 4.46C, average life expectancy has the highest positive correlation of 0.46 in PC 1 while first wave deterministic R_o (first wave D) has the highest positive correlation in PC 2, equal to 0.47. The two socio-economic variables explaining the most clustering are the average life expectancy (LE) and the stringency index (SI).

4.4.8 Multivariate analysis, clustering and prediction for OECD countries

The multivariate analysis using ordinary least square method, clustering analysis using PCA, K-means and Hierarchy clustering and the prediction of the percentage GDP devoted to health expenditure using MLP regressor which is a deep learning method, Linear, Lasso and Ridge regression for OECD countries is presented.

- Figure 4.48 corresponds to the ordinary multivariate least square method with $R^2 = 0.90$.

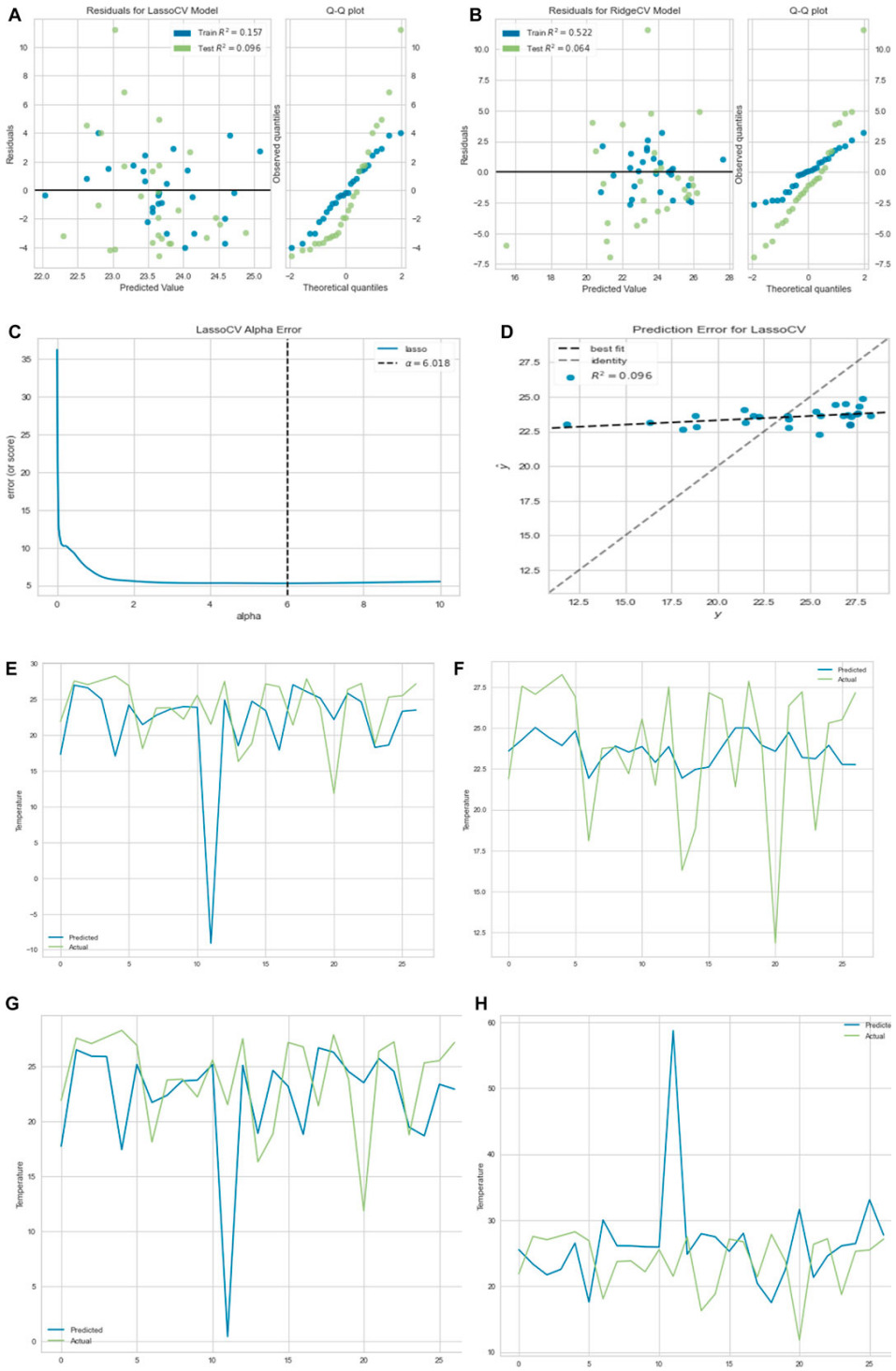


Figure 4.45: (A) Residual plot for lasso regression. (B) Residual plot for ridge regression. (C) Lasso regression cross-validation error. (D) Prediction error for lasso regression curve. (E) Linear regression prediction plot. (F) Lasso regression prediction plot. (G) Ridge regression prediction plot. (H) MLP regression prediction plot.

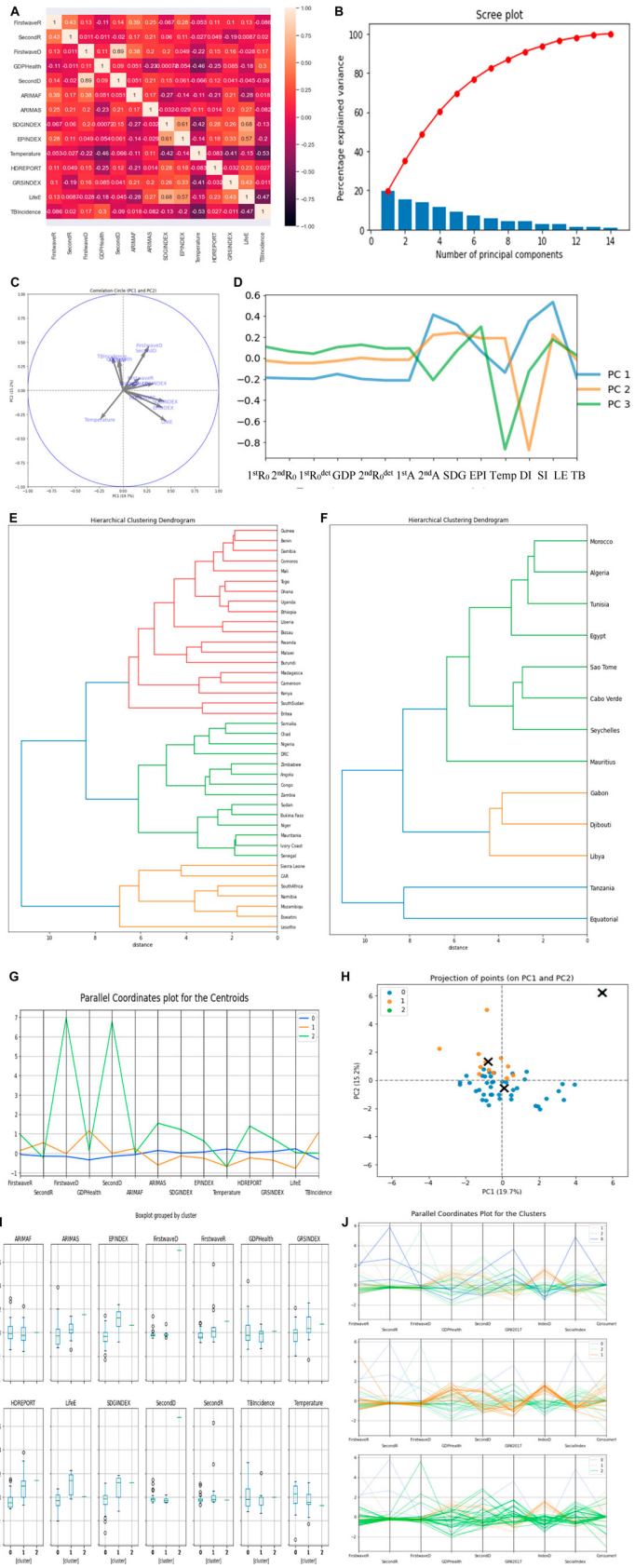


Figure 4.46: (A) Heatmap of the parameter's correlations. (B) Scree plot. (C,D) PCs visualization. (E,F) Hierarchy clustering dendrogram. (G) Parallel coordinates plot for the centroids. (H) Projection points for PC1 and PC2. (I) Boxplot of the clusters. (J) Parallel coordinates plot for the clusters.

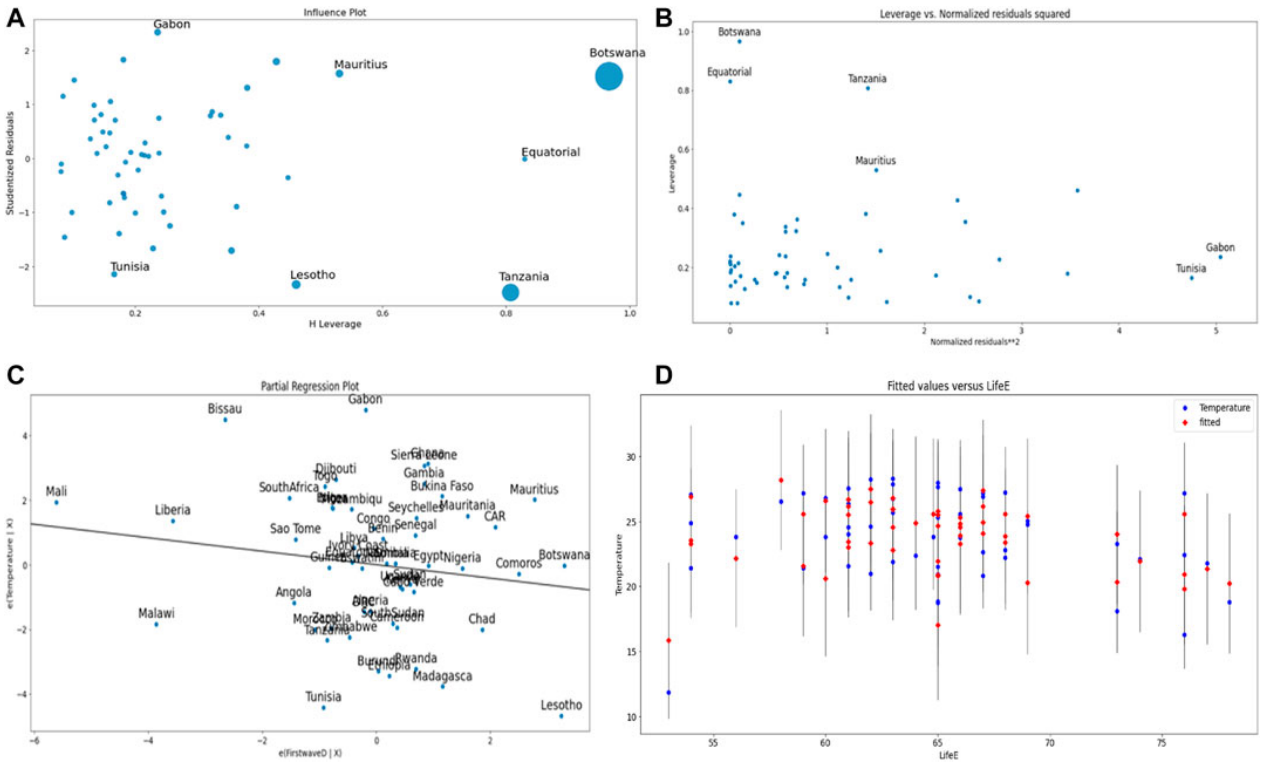


Figure 4.47: (A) Influence plot. (B) Leverage vs. normalized residuals squared plot. (C) Partial regression plot of the temperature vs. maximal R_0 of the first wave. (D) Fit plot. The first wave maximal R_0 is negatively correlated with the mean temperature of the country ([101, 102]).

- Figure 4.48A shows Iceland, United States, Austria, and Belgium as outliers not fitting the data.
- The partial regression plots in Figure 4.48C shows that the best correlation observed is between CHE/GDP and the demo-economic index DI as observed before in [232].
- Cross-validation method was used to choose the best parameter α for the modeling as shown in Figure 4.49D.
- For ridge regression, $\alpha = 0.005$ with a mean square error of 1.905, and for Lasso regression, $\alpha = 0.027$ with a mean square error = 1.657.
- For Figure 4.49E, training score = 0.993 and test score = 0.535.
- For Figure 4.49F, training score = 0.898 and test score = 0.629.
- For Figure 4.49G, training score = 0.983 and test score = 0.259.
- For Figure 4.49H, training score = -0.072 and test score = -0.196.

It is evident from these results that the lasso regression best predicts percentage of GDP devoted to health expenditure with the highest test score and predicted values are very close. All the regression methods give about the same result with the maximum accuracy for the ridge regression.

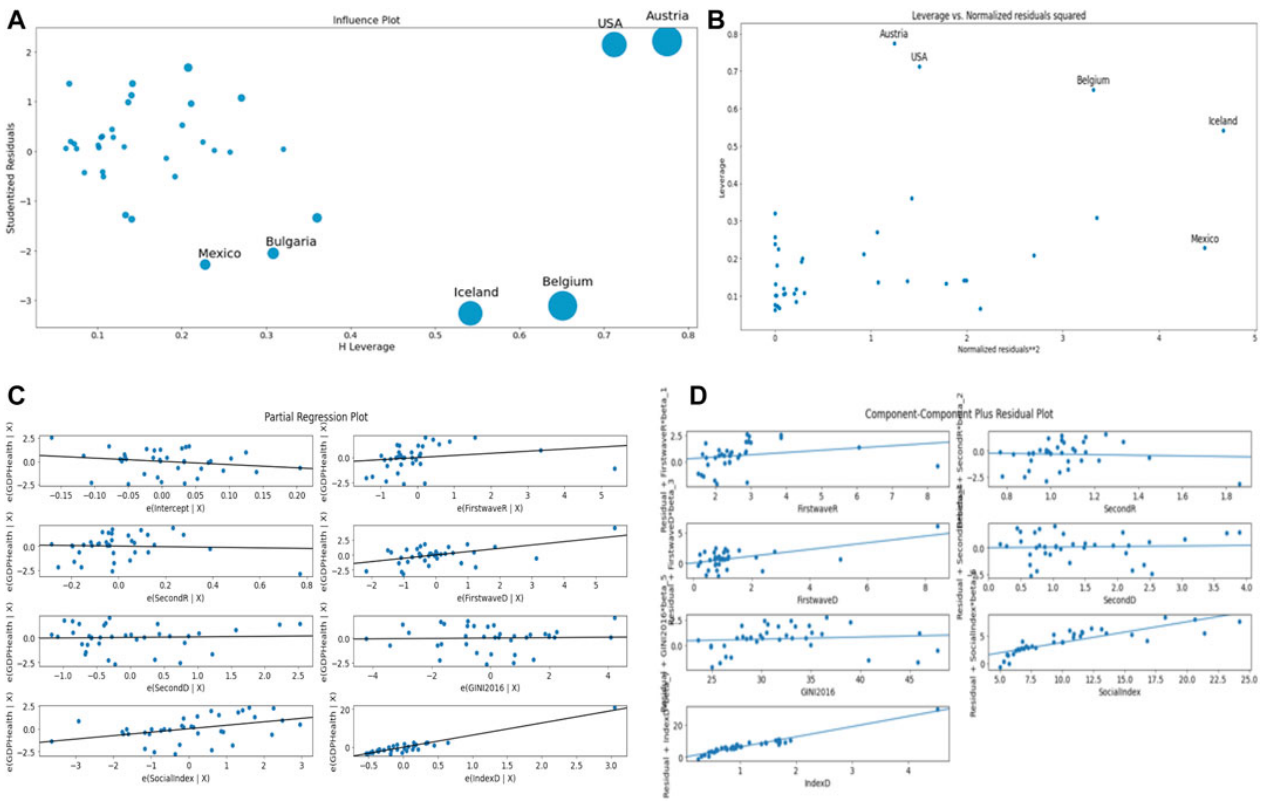


Figure 4.48: (A) Influence plot. (B) Leverage vs. Normalized residuals squared plot. (C) Partial regression plot. (D) Component–Component plus residual plot.

In Figure 4.50E and Figure 4.50F, the first cluster has 20 countries, and the second has 5 countries, which are United States and Bulgaria on the same hierarchy, Mexico and Costa Rica on the same hierarchy, and Chile standing alone. The third cluster has 12 countries. Only the two highest cluster dendrograms is presented. In Figure 4.50C, the Gini index and social fracture index have the highest positive correlation of 0.45 and 0.46, respectively, in PC 1 while the percentage of GDP devoted to health expenditure and demo-economic index have the highest positive correlation in PC 2, whose values equal to 0.65 and 0.41, respectively. The two main clusters correspond both to developed countries, but in the first, countries are more continental, and in the second, countries are more maritime, which could be explained by their difference in consumer confidence index (CCI), which is less important in maritime countries than in continental ones.

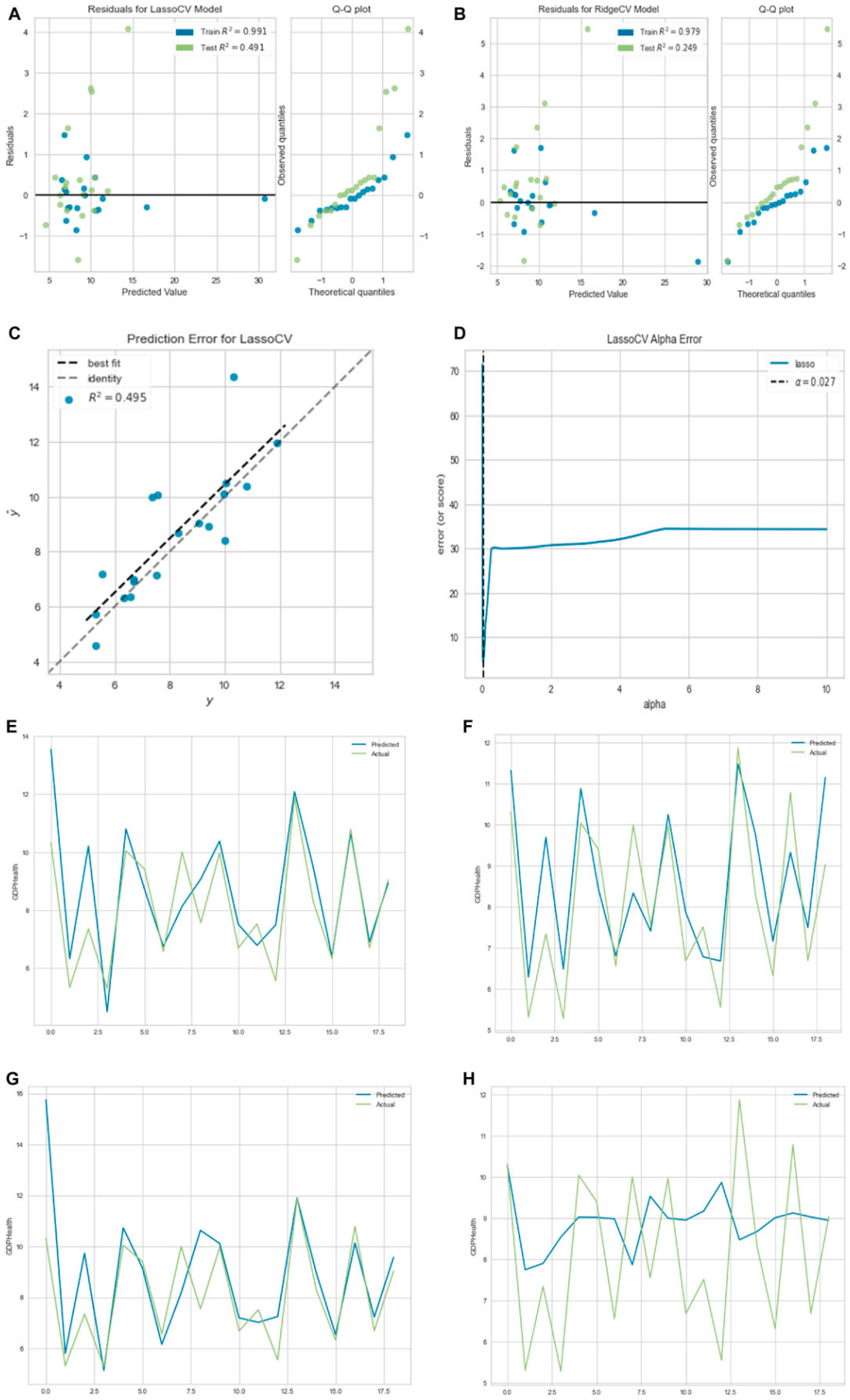


Figure 4.49: (A) Residual plot for lasso regression. (B) Residual plot for ridge regression. (C) Prediction error for lasso regression curve. (D) Lasso regression cross-validation error. (E) Linear regression prediction plot. (F) Lasso regression prediction plot. (G) Ridge regression prediction plot. (H) MLP regression prediction plot.

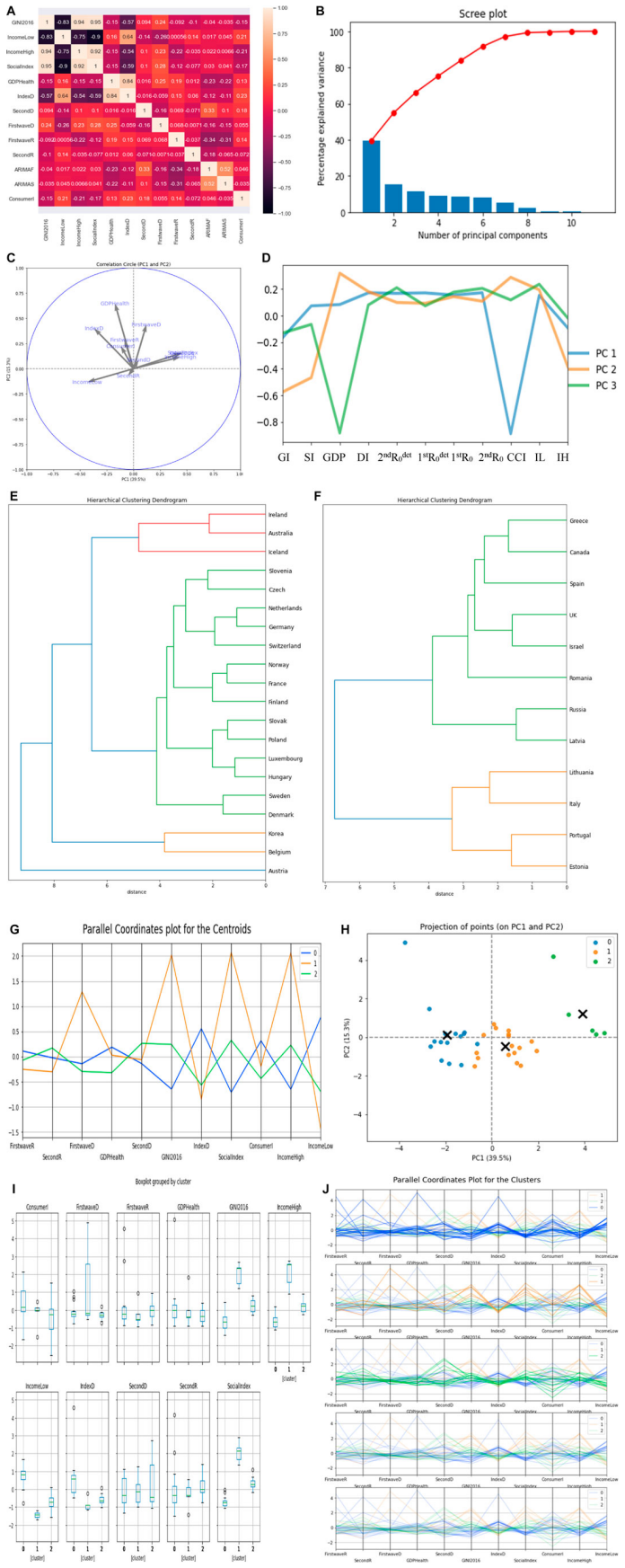


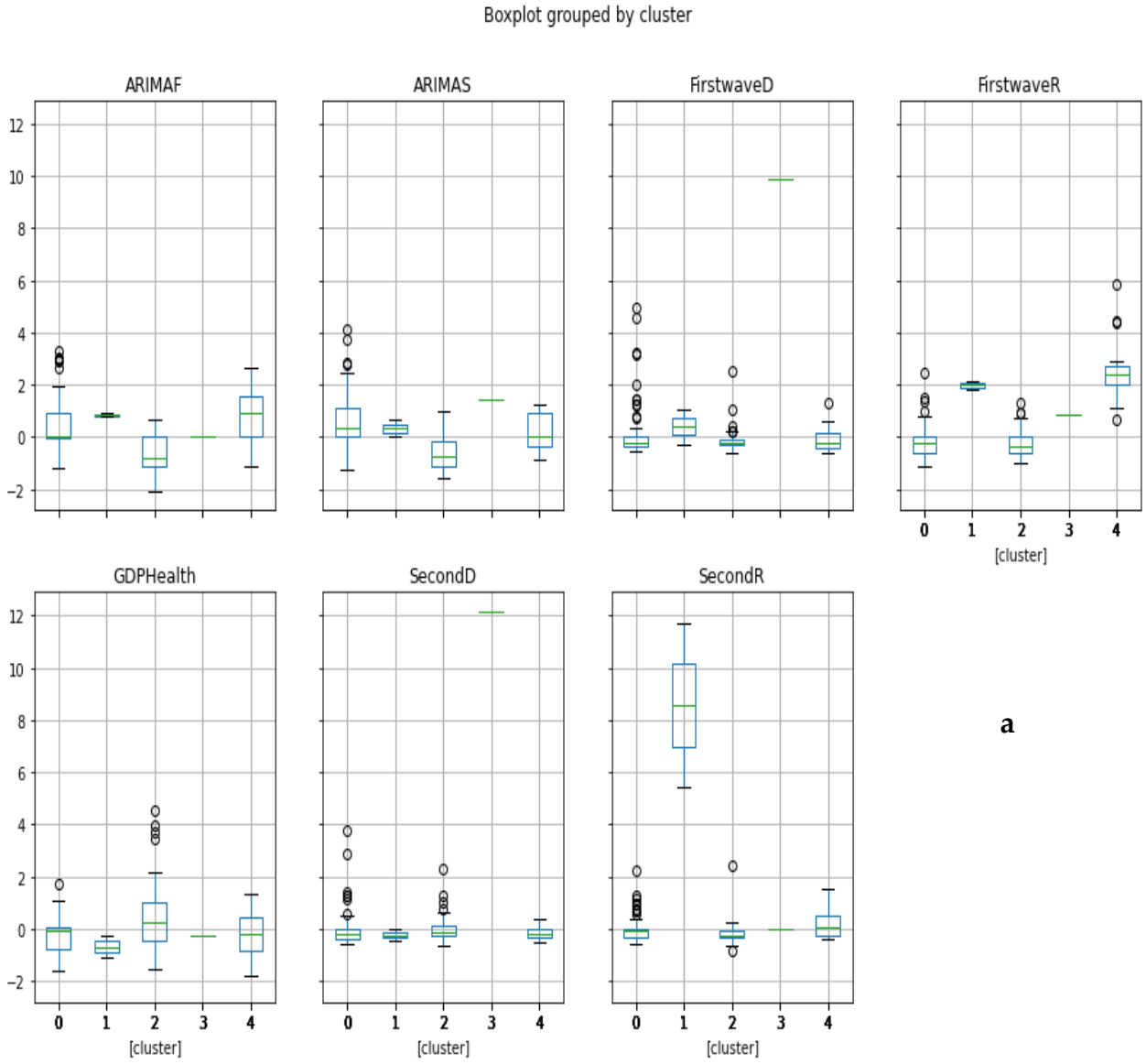
Figure 4.50: (A) Heatmap of the parameter's correlations. (B) Scree plot. (C,D) PCs visualization. (E,F) Hierarchy clustering dendrogram. (G) Parallel coordinates plot for the centroids. (H) Projection points for PC1 and PC2. (I) Boxplot of the clusters. (J) Parallel coordinates plot for the clusters.

Table 4.10: OECD countries data.

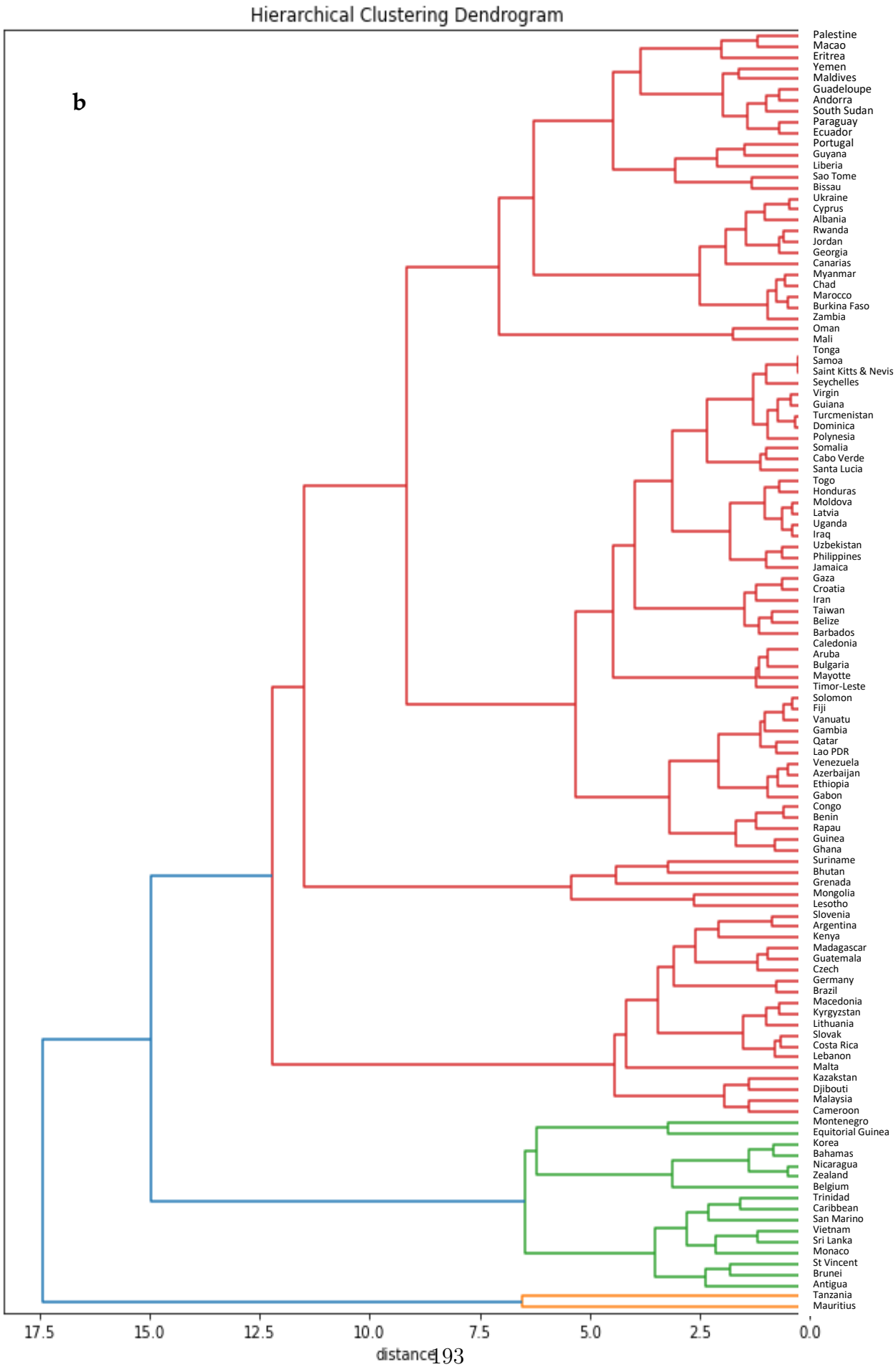
S/N	Country Name	10% LI	10% HI	SF index	CCI	Gini index	Demographic Index
1	AUSTRALIA	2.8	26.1	9.32	100.86	32.5	1.00
2	AUSTRIA	3.3	22.5	6.82	99.15	28.0	4.52
3	BELGIUM	3.6	20.6	5.72	100.75	25.8	1.80
4	BULGARIA	1.9	31.9	16.79	-	40.8	0.44
5	CANADA	2.6	24.2	9.31	-	30.3	1.15
6	CHILE	1.8	37.1	20.61	96.55	46.0	0.44
7	COSTA RICA	1.5	36.3	24.20	98.94	47.8	0.31
8	CZECH	4.0	22.2	5.55	98.97	24.9	1.38
9	DENMARK	4.0	21.2	5.30	100.50	26.4	1.90
10	ESTONIA	2.3	26.3	11.43	97.59	30.5	0.59
11	FINLAND	4.0	21.2	5.30	100.27	26.9	1.71
12	FRANCE	3.5	24.2	6.91	98.12	30.1	1.63
13	GERMANY	3.5	23.5	6.71	98.97	28.9	1.70
14	GREECE	2.2	25.4	11.55	98.11	30.6	0.67
15	HUNGARY	3.1	22.5	7.26	98.88	28.9	0.92
16	ICELAND	4.1	20.6	5.02	-	25.0	1.69
17	IRELAND	3.1	24.4	7.87	100.48	29.5	0.88
18	ISRAEL	2.0	26.3	13.15	99.15	34.8	0.57
19	ITALY	2.1	24.4	11.62	100.25	33.4	0.75
20	KOREA	2.2	22.0	10.00	99.43	34.5	0.76
21	LATVIA	2.4	26.1	10.88	95.05	35.1	0.57
22	LITHUANIA	2.1	28.4	13.52	100.98	36.1	0.49
23	LUXEMBOURG	3.5	22.1	6.31	98.52	31.8	0.84
24	MEXICO	1.7	36.4	21.41	99.42	45.8	0.25
25	NETHERLANDS	3.3	22.7	6.88	99.93	28.5	1.45
26	NORWAY	3.4	20.6	6.06	-	26.2	1.66
27	POLAND	3.1	23.4	7.55	98.70	28.1	0.84
28	PORTUGAL	2.4	26.2	10.92	97.41	31.7	0.86
29	ROMANIA	1.6	24.9	15.56	-	35.0	0.36
30	RUSSIA	2.9	29.1	10.03	96.12	33.1	0.53
31	SLOVAK	3.3	21.6	6.55	97.71	23.6	1.02
32	SLOVENIA	3.6	20.4	5.67	96.34	24.9	1.46
33	SPAIN	2.0	24.7	12.35	96.61	33.0	0.73
34	SWEDEN	3.5	22.6	6.46	101.89	27.5	1.69
35	SWITZERLAND	3.4	24.1	7.09	97.47	29.9	1.68
36	UK	2.7	28.6	10.59	98.71	36.6	0.94
37	USA	1.6	29.2	18.25	98.68	39.0	0.93

4.4.9 Multivariate analysis, clustering and prediction for all countries in the world

Figure 4.52 show a positive correlation between the slope of the logarithmic regression curve of the new cases of COVID-19 and the economic index CHE/GDP and the correlation coefficient for the first wave remains for all countries close to that for developed countries. The result for the linear regression is as follows:
slope = 0.01214439, intercept = −0.0159087, r-value = 0.3655, p-value = 2.71 $\times 10^{-6}$, standard error = 0.00249223, R-squared = 0.13359, RMSE = 0.0819603345.



a


Hierarchical Clustering Dendrogram
b
Palestine
Macao
Eritrea
Yemen
Maldives
Guadeloupe
Andorra
South Sudan
Paraguay
Ecuador
Portugal
Guyana
Liberia
Sao Tome
Bissau
Ukraine
Cyprus
Albania
Rwanda
Jordan
Georgia
Canarias
Myanmar
Chad
Marocco
Burkina Faso
Zambia
Oman
Mali
Tonga
Samoa
Saint Kitts & Nevis
Seychelles
Virgin
Guiana
Turcmenistan
Dominica
Polynesia
Somalia
Cabo Verde
Santa Lucia
Togo
Honduras
Moldova
Latvia
Uganda
Iraq
Uzbekistan
Philippines
Jamaica
Gaza
Croatia
Iran
Taiwan
Belize
Barbados
Caledonia
Aruba
Bulgaria
Mayotte
Timor-Leste
Solomon
Fiji
Vanuatu
Gambia
Qatar
Lao PDR
Venezuela
Azerbaijan
Ethiopia
Gabon
Congo
Benin
Rapau
Guinea
Ghana
Suriname
Bhutan
Grenada
Mongolia
Lesotho
Slovenia
Argentina
Kenya
Madagascar
Guatemala
Czech
Germany
Brazil
Macedonia
Kyrgyzstan
Lithuania
Slovak
Costa Rica
Lebanon
Malta
Kazakstan
Djibouti
Malaysia
Cameroon
Montenegro
Equitorial Guinea
Korea
Bahamas
Nicaragua
Zealand
Belgium
Trinidad
Caribbean
San Marino
Vietnam
Sri Lanka
Monaco
St Vincent
Brunei
Antigua
Tanzania
Mauritius
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
distance

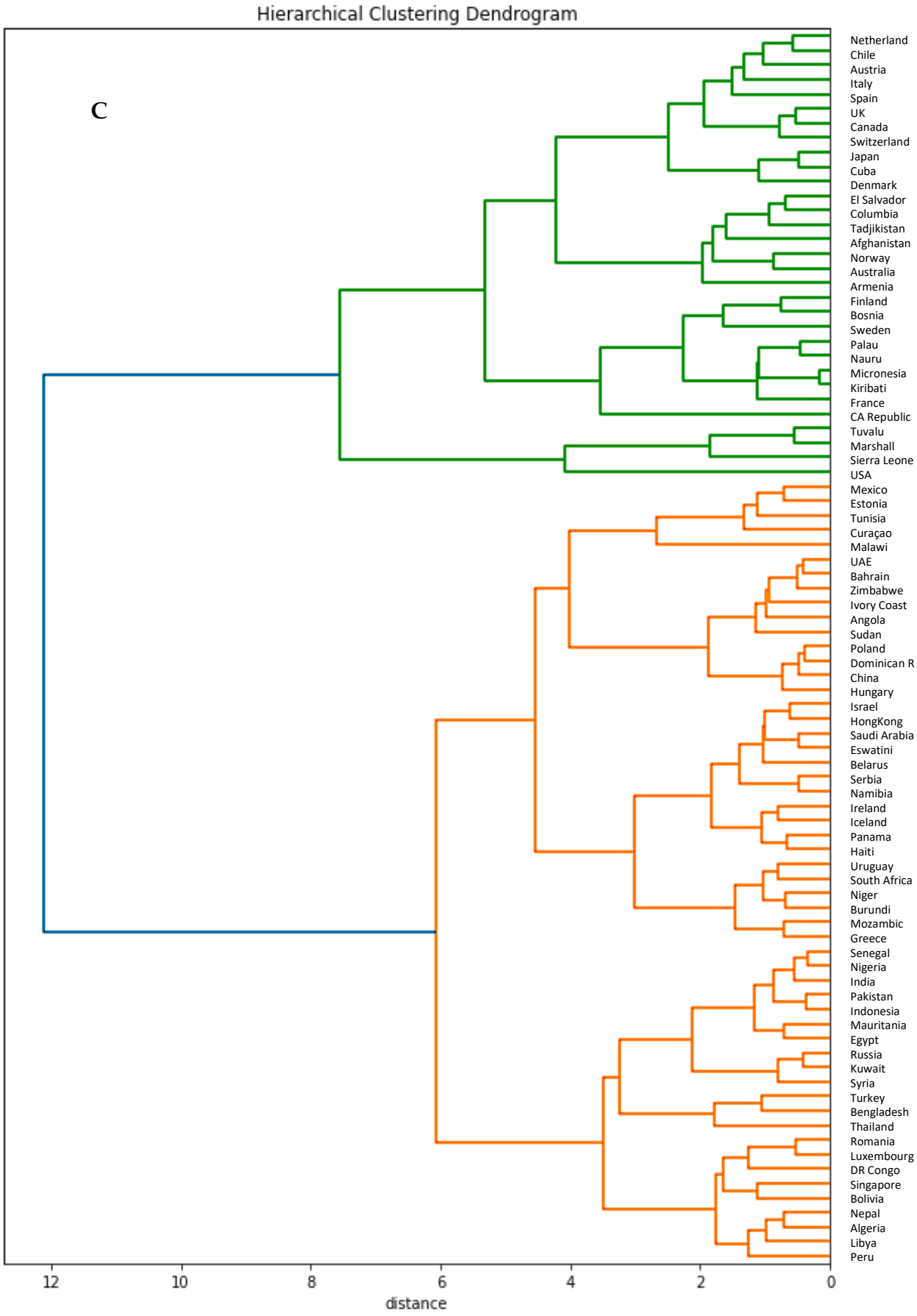


Figure 4.51: (**a**) Boxplots of the clusters. Visualizations of (**b**) more "developing" (in red with some notable exceptions such as the Czech Republic and Germany) and (**c**) more "developed" (in green and partially in orange) countries parts of the hierarchy tree.

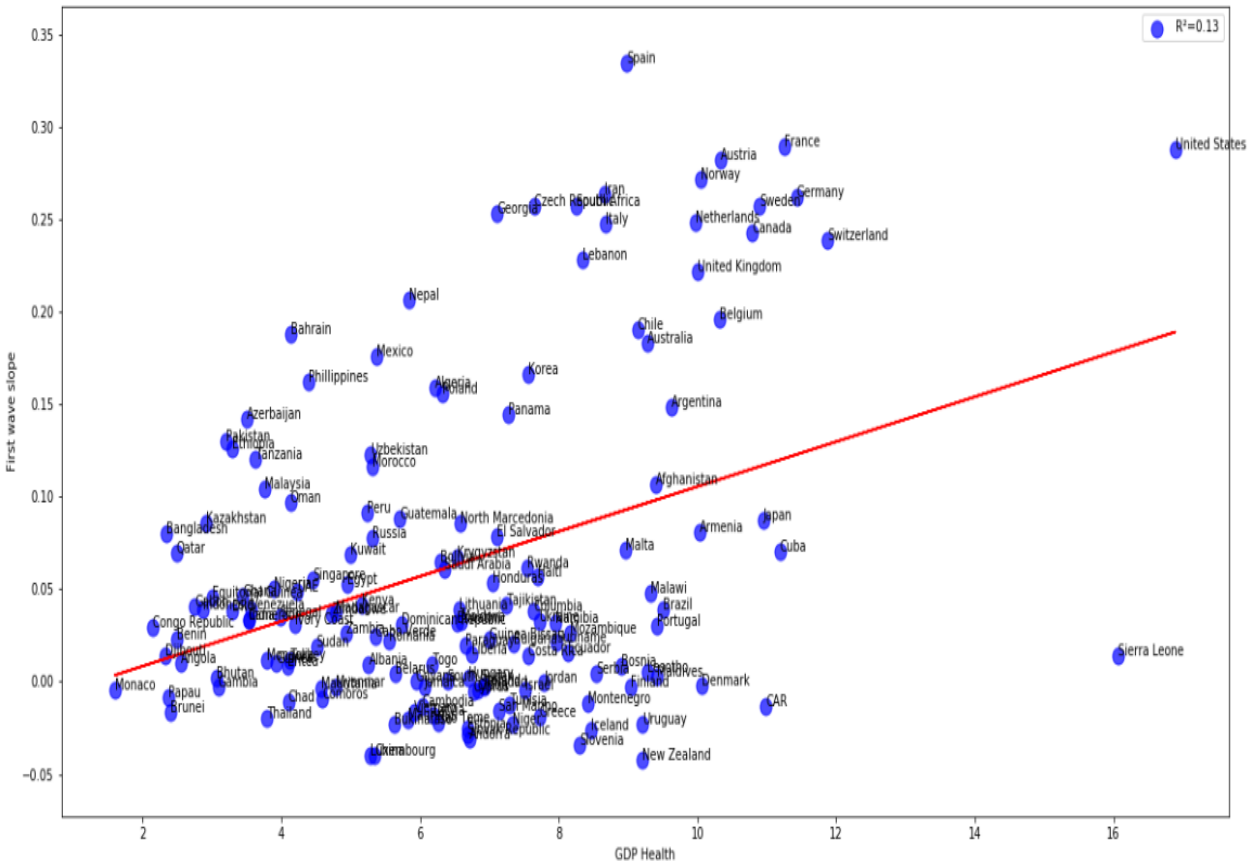


Figure 4.52: Regression plot of first wave exponential regression slope vs CHE/GDP for all countries.

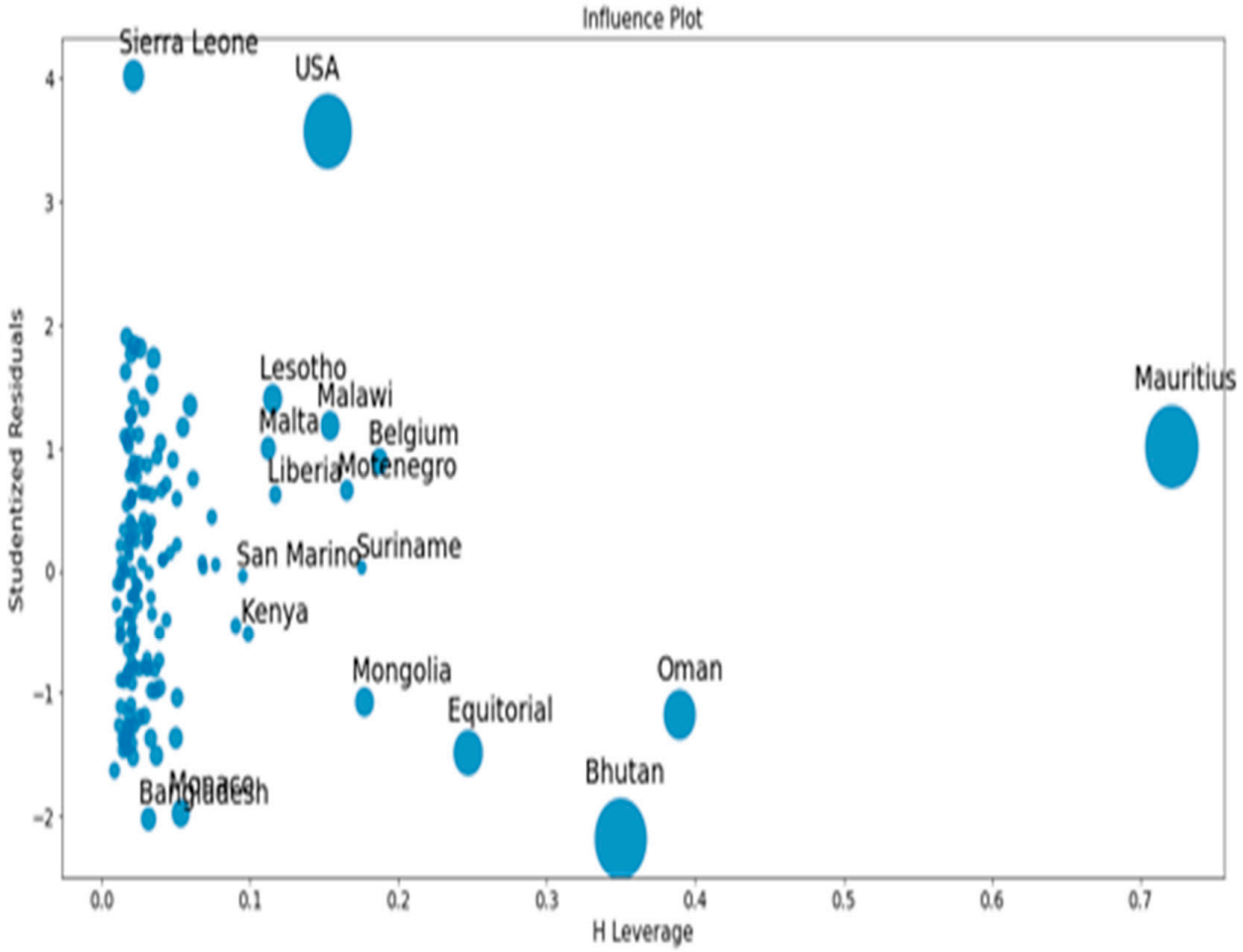
The clustering and multivariate analysis using ordinary least square method for all countries from epidemic and economic variables is described in Figure 4.51 to Figure 4.54 and shows several features:

1. The hierarchical clustering (Figure 4.51 and Figure 4.54(b)) shows a trend common to developed countries (shown in green), with the notable exception of Germany and Czechia,

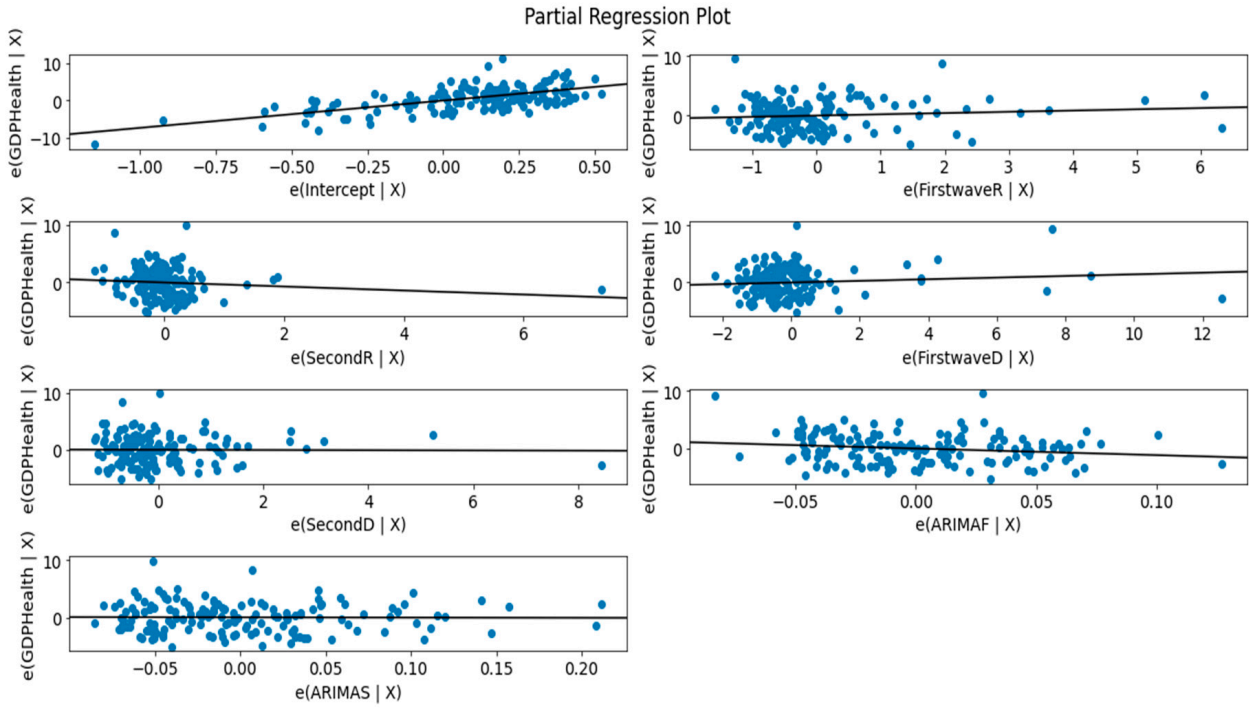
2. The principal component analysis shows the importance of the CHE/GDP index in the first principal component (Figure 4.54(a),(c),(d)) and of the deterministic $R_0(R_0^{det})$ of the exponential phase of the first wave in the second principal component and of the second wave in third principal component (Figure 4.54(e));

3. The analysis of parallel coordinates for cluster centroids also shows the importance of the deterministic R_0 in the discrimination of clusters (Figure 4.55);

4. The analysis of the residuals shows a good explanatory power of the first three principal components (60% of the total variance in Figure 4.54(c), confirmed by the projections on the two first principal planes of Figure 4.54(d)-(e)), and a weak correlation of the principal components with these residuals (Figure 4.53(a)-(b)).

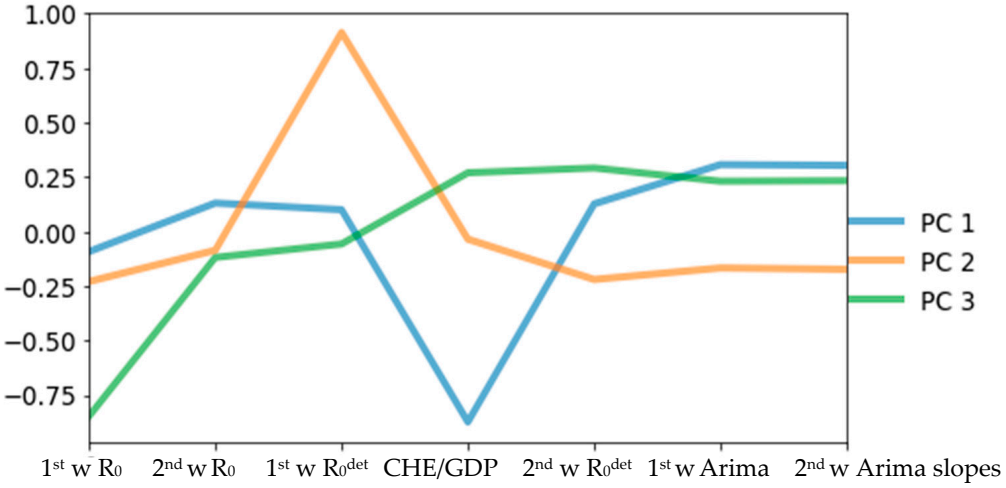


(**a**)

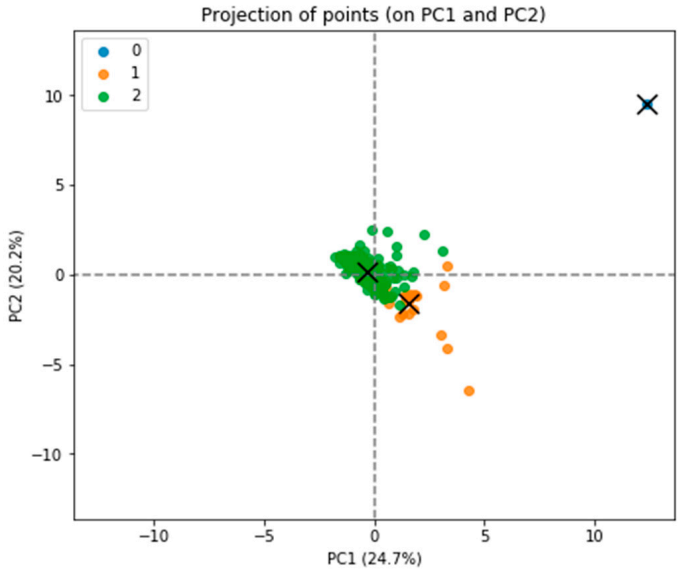


(**b**)

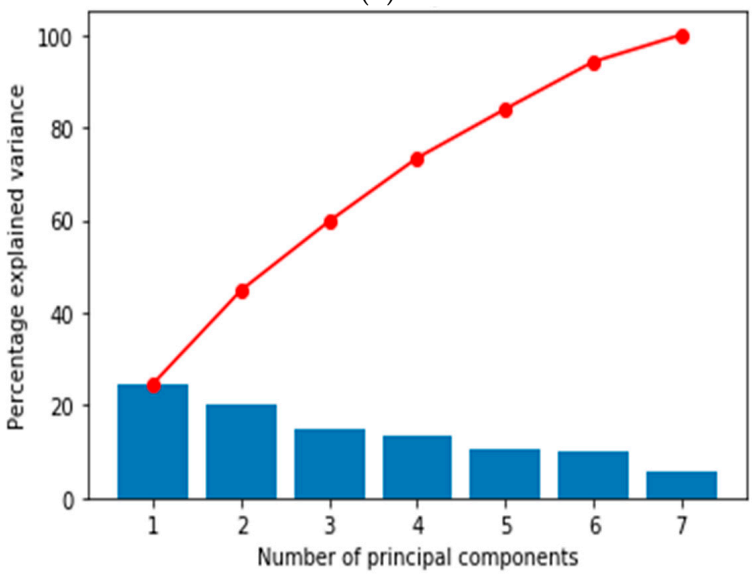
Figure 4.53: (**a**) Leverage vs normalized squared residuals plot. (**b**) Residuals regression plots for initial variables.


1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
PC 1
PC 2
PC 3
1st w R0
2nd w R0
1st w R0det
CHE/GDP
2nd w R0det
1st w Arima
2nd w Arima slopes

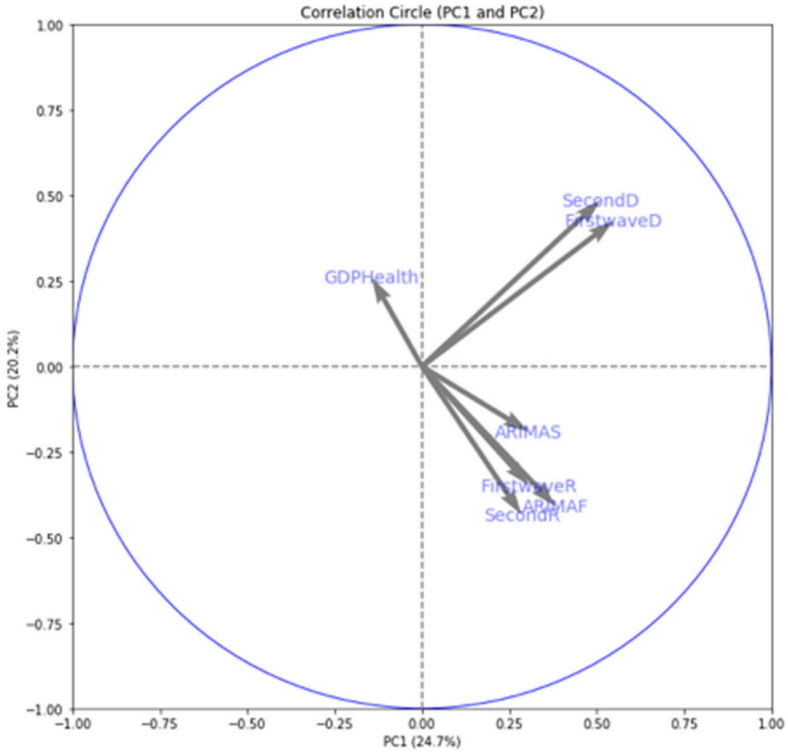
(**a**)


Projection of points (on PC1 and PC2)
0
1
2
PC2 (20.2%)
PC1 (24.7%)

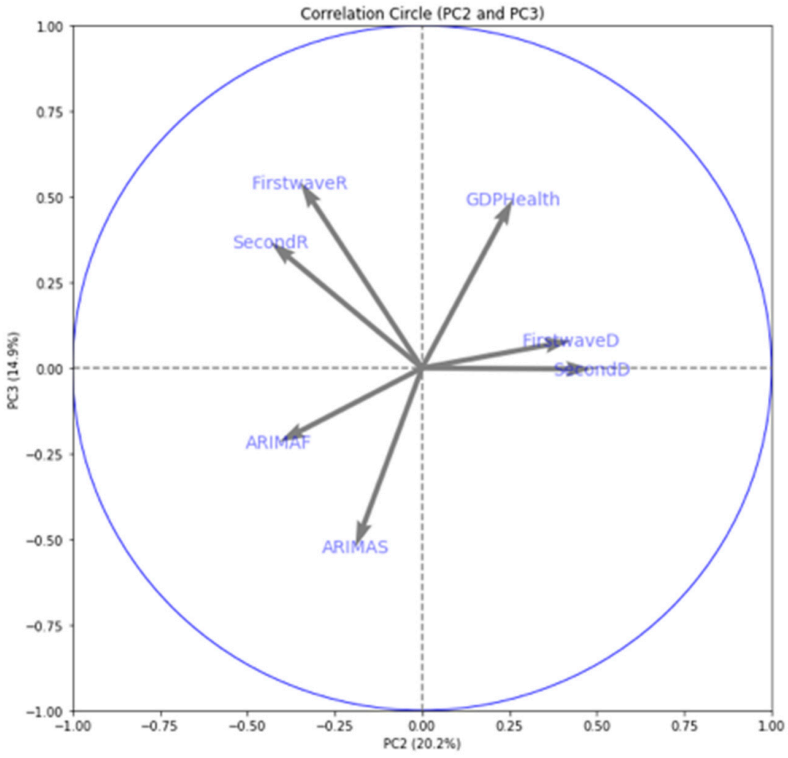
(**b**)


Percentage explained variance
Number of principal components

(**c**)



(**d**)



(**e**)

Figure 4.54: (**a**) Principal components (PC) plot from the principal component analysis (PCA) on the initial variables: first and second waves maximum R_0, 1st wave R_0 and 2nd wave R_0, deterministic R_0, 1st wR_0^{det} and 2nd wR_0^{det}, 1st wave Arima slope, 2nd wave Arima slope, and CHE/GDP. (**b**) Projection of the points corresponding to 204 countries of the PCA's plot on the first PC plane with more developed countries in green and more developing in orange. (**c**) Explained variance plot. (**d**,**e**) Correlation circles for the two first PC planes.

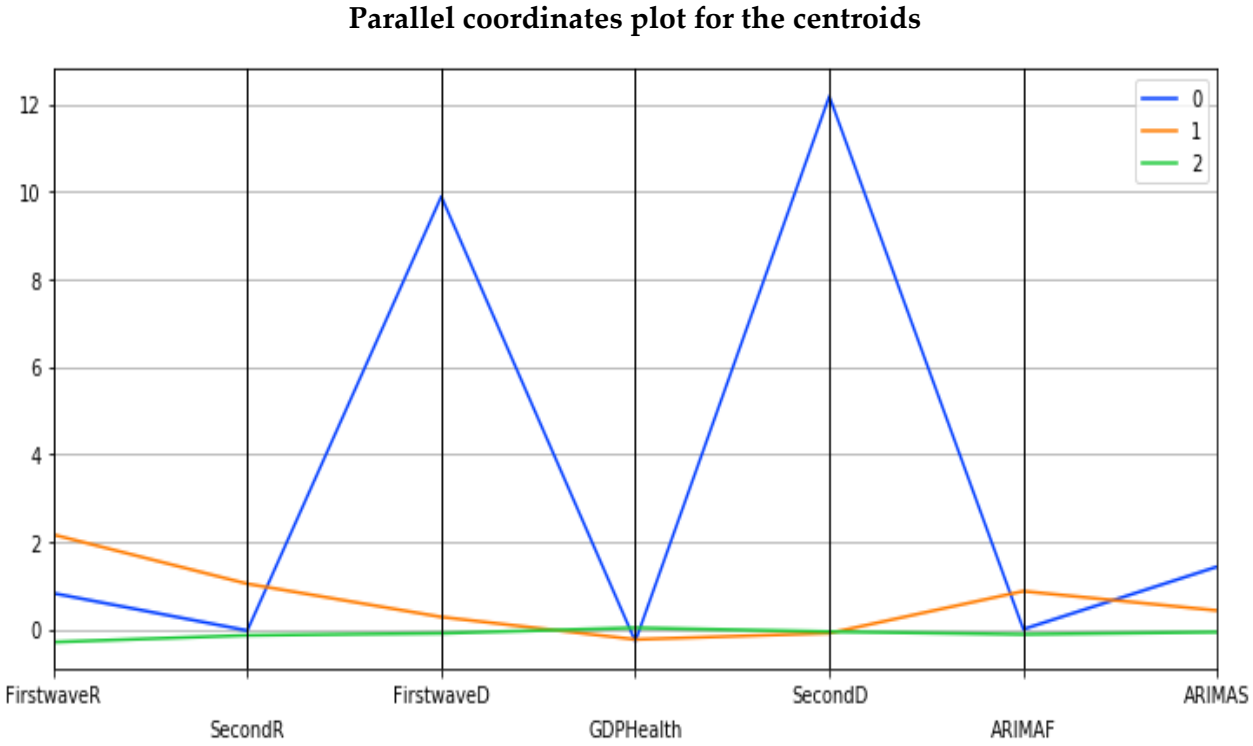


Figure 4.55: Parallel coordinates for cluster centroids.

The multivariate analysis using ordinary least square method, clustering analysis using PCA, K-means and Hierarchy clustering and the prediction of the percentage GDP devoted to health expenditure using MLP regressor which is a deep learning method, linear, lasso and ridge regression for all countries is presented.
To conclude this section, a new approach to the modeling of epidemiology, socio-economic, geo-climatic, and demographic determinants of COVID-19 outbreak during the exponential phase has been developed. Some of these determinants have shown high correlation with epidemiologic parameters as it can be seen in the heatmap diagrams in Figure 4.29G, Figure 4.41A, Figure 4.46A, and Figure 4.50A, explaining the role of each variable thanks to these correlations.
For developed and developing countries, the lasso regression reduced the correlation between the social fracture index and the 10% highest income, while for OECD countries, the correlation between the Gini index and social fracture index was reduced to zero. Some of the variables were not used in the optimization method-OLS due to multicollinearity observed on results summary. For the two sets of countries, consumer confidence index, opposite of the initial autocorrelation slope averaged on 6 days for the first and second wave, 10% lowest income, and 10% highest income were not used in the modeling. The R-squared for OLS results for developed, developing, and OECD countries are 0.76 and 0.90, respectively, which shows a high significance rate (Figure 4.41E, Figure 4.41F and Figure 4.48).
The principal component analysis shows high correlation for the numbers of new cases used in this analysis. The social fracture index has high correlation in PC1 for both cases, while in PC2, percentage of GDP devoted to health expenditure was dominant for OECD countries, and maximum R_o for the first wave was dominant for both developed and developing countries (see Figure 4.41C and Figure 4.50C). What can be deduce from all these observations is that the socio-economic determinants are key to the modeling of infectious diseases like COVID-

19 as these parameters give high signals on the trend during the spread of the pandemic for various countries ([138, 139, 143, 233]).
Similarly, there are a lot of differences between the first and second wave results concerning the exponential regression slope and the auto-correlation initial slope: while some countries have higher figures for the first wave, others have lower figures for the second wave and vice versa. This was also evident for the regression plot where some countries have negative correlation values for the first wave of some growth parameters with the CHE/GDP and positive for the second wave, and vice versa for other countries. These phenomena prove that the way the pandemic spread in the second wave is different from what was experienced in the first wave. In the principal component analysis, we discovered that first wave deterministic R_0 and CHE/GDP health had high weights in first and second principal components (PC1 and PC2), which are dominant components in the PC analysis.
More precisely, on Figure 4.42, for the first wave the opposite of the initial auto-correlation slope is positively (resp. negatively) correlated with the CHE/GDP (resp. maximum R_0) for developed countries, which could correspond to the efficiency of the mitigation measures decided in these countries. This is confirmed on Figure 4.44, where the first wave exponential regression slope is positively correlated with the CHE/GDP in a mix of developed and developing countries. The Figure 4.43a shows the same type of effect of public health policies in developed countries for the first wave, where CHE/GDP increases with the first wave exponential regression slope, but this result is inverted on Figure 4.43b for the second wave perhaps due to a rationalization of the care activity between the first two waves. For the first wave of all countries, Figure 4.52 shows the same positive correlation as Figure 4.43a between the exponential regression slope and CHE/GDP.
Clustering of all countries is then studied on Figure 4.51, Figure 4.55 and Figure 4.54. Figure 4.51a shows the box-plot of the seven initial variables used in hierarchical clustering: the first and second wave opposite of the initial auto-correlation slope (respectively ARIMAF and ARIMAS), exponential regression slope and maximum R_0 (respectively FirstwaveD, SecondD, FirstwaveR, SecondR), and CHE/GDP. The box-plots contain five clusters represented in Figure 4.51(b)-(c) corresponding to more "developing" (in red with some notable exceptions such as the Czech Republic and Germany) and (c) more "developed" (in green and partially in orange) countries parts of the hierarchy tree, with a small "exotic" cluster for Tanzania and Mauritius. Figure 4.54(a)-(e) shows the results of the principal component analysis (PCA), with (a) the three principal components declined on the initial variables calculated for all countries (first and second waves maximum R_0's denoted first wR_0 and second wR_0, deterministic R_0's denoted first wR_0^{det} and second wR_0^{det}, Arima slopes denoted first wArima, second wArima slopes, and the current health expenditure as gross domestic product percentage denoted CHE/GDP), (b) the projection of the points corresponding to countries of the PCA's plot on the first PC plane, (c) the explained variance plot and (d,e) the correlation circles for the first three principal components with projection of the initial variables as vectors (having 195 components corresponding to the 195 countries of the table in appendix on the corresponding principal planes. In Figure 4.54(a), the main initial variable in the linear combination of the first (resp. the second) principal component is the first wave deterministic R_0^{det} (resp. the CHE/GDP) and these two initial variables R_0^{det} and CHE/GDP are anti-correlated as we have already noticed when commenting before on the Figure 4.42 (a country devoting a large share of its GDP to health expenditure reduces the occurrence of new cases). Figure 4.54(b) gives the projection of 204 countries on the first PC plane and distinguishes two main clusters of 118 and 85 countries, respectively, plus a singleton repre-

senting Botswana, with more developed countries in green and more developing countries in orange. Figure 4.54(c) shows that 60% of the variance is explained by the three first PCs, and Figure 4.54(d)-(e) presents the correlation circles with projection of the initial variables as vectors on the corresponding two principal planes (PC1, PC2) and (PC2, PC3), showing such as in Figure 4.54(a) the preeminence of the opposite vectors, the first wave deterministic R_0 and the CHE/GDP. Figure 4.55 shows also for the first K-means cluster the importance of the first wave deterministic R_0.
Finally, Figure 4.53(a)-(b) corresponds to the ordinary multivariate least square method. Figure 4.53(a) shows the eccentric position of developed countries such as Belgium and USA and developing countries such as Equatorial Guinea and Suriname as outliers not fitting the data bulk, and Figure 4.53(b) the concentration of the initial variable CHE/GDP with the first and second waves deterministic R_0^{det}, in agreement with the fact that they are the most dominant initial variables in PCA and k-means clustering.

4.5 Functional Data Analysis: Application to Daily Observation of COVID-19 Prevalence

The main purpose of this section is to revisit the data on COVID-19 from public databases using methods that are still little used, such as Functional Data Analysis (FDA), in which there is a great deal of theoretical work [234, 235, 236, 237, 238, 239, 240, 241, 242, 243], but practical applications are still rare. In particular, the generalization at the FDA of classical finite-dimensional methods, such as estimation, regression and principal component analysis, shows that it is possible to process epidemic data obtained from a large sample (approximately one million data) concerning the incidence, mortality and exogenous or endogenous factors associated with the COVID-19 epidemic. This sampling concerns about ten variables (sometimes missing, in particular those concerning risk factors) and includes longitudinal (about 600 days all over the world but in this case about 469 days for French departments) and cross-sectional data (about 200 countries, some comprising up to a hundred documented regions while for France which is the main focus of this section there are 101 departments), for the monitoring and prediction of a pandemic whose origin and end are still uncertain, but which, by its magnitude and its dramatic consequences (around 5 million deaths) justifies such a descriptive statistical investment. The descriptive study of many factors associated with the epidemic (namely the socio-economic and geoclimatic ones) makes it possible to understand closely linked mechanisms, those exogenous or endogenous to the viral pathogeny of propagation and endogenous ones of pathogenicity [101, 102, 103, 139, 233]. The seminal paper by Bernoulli [1] on the SI model is the origin of all future discussions about epidemic modelling, in particular the first by d'Alembert [70] in his Opuscules mathématiques and Lambert until its last recent refinements. It contains the explicit solution of the SI model given in Equation (3.36) and permits to consider the empiric $S(t)$ and $I(t)$ curves as functions. It can be manipulated (after smoothing) as elements in a functional space. The goal here is to apply to these elements the FDA descriptive techniques like Canonical Correlation Analysis (CCA), K-means clustering and Functional Principal Component Analysis (FPCA) in order to compare these functions among different departments in France before vaccination started and after vaccination has started and to finally make some predictions about the evolution of the disease in France. The functional variables are numbers of ICU cases, daily deceased (DD), daily return home (DRH) and hospitalization which are given as

X_1, X_2, X_3, X_4 while response variables are given as $Y_1, Y_2, Y_3, Y_4, Y_5, Y_6$ are numbers of recovery, deaths, infected, vaccination, vaccination per 1000 population and number of tests. All the data are from [119, 244, 245]. This approach is complementary of the works estimating the parameters of the functions S and I (namely R_0 and $\mathfrak{B}$) presented in Section 3.3.

4.5.1 Time series and curve fitting

The time series analysis of daily new cases in France, daily hospitalization in three French departments out of 101 in France, and curve fitting for two French departments is presented.

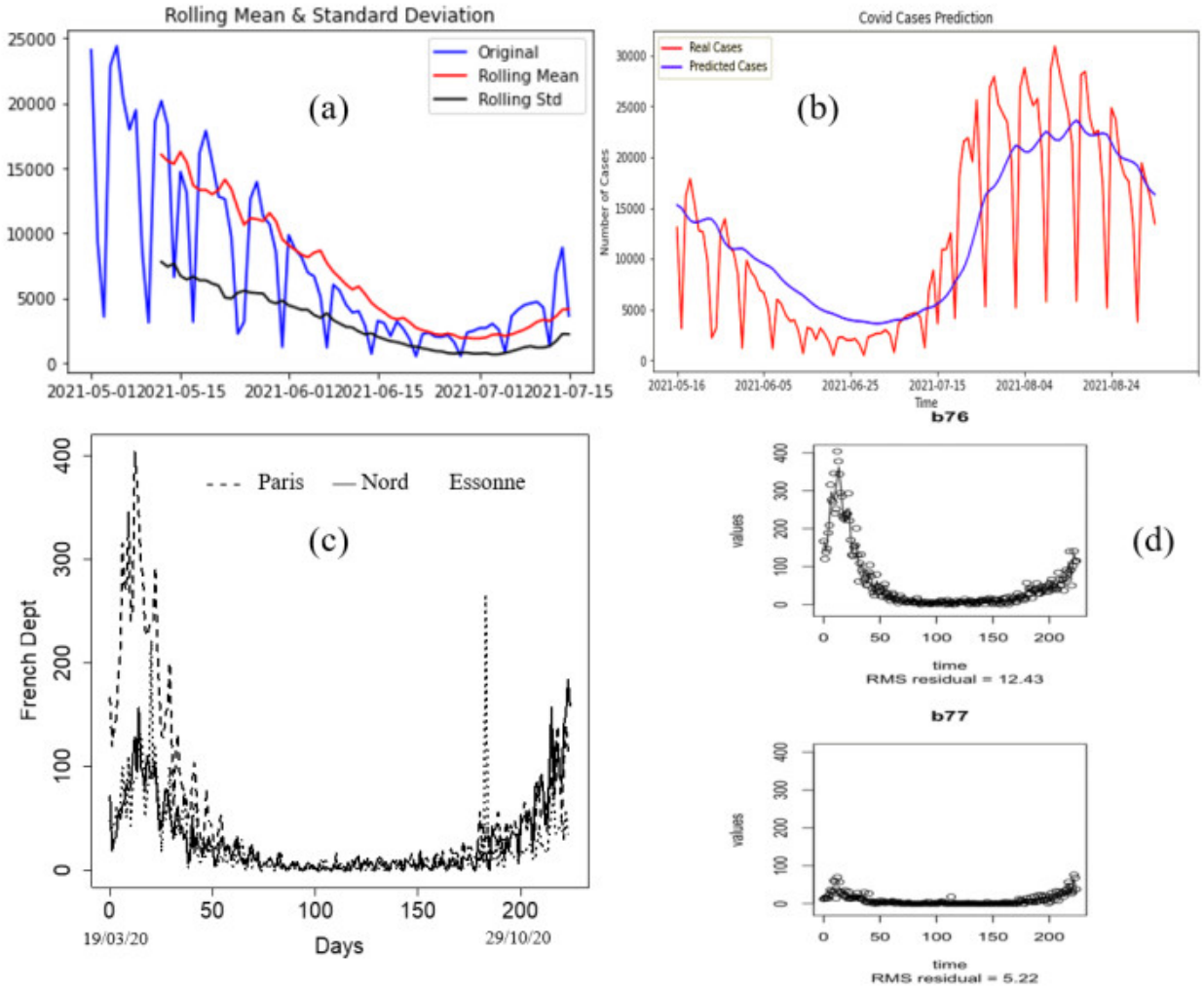


Figure 4.56: (a) Time series modelling of daily new cases between 01/05/2021–15/07/2021 in France. (b) GRU deep learning forecasting method for daily new cases between 25/02/2020–03/09/2021 in France. (c) Daily hospitalization cases in three French departments: Nord, Paris and Essonne. (d) Fit curve for hospitalization cases in Paris and Seine-Maritime.

The RMSE for other curve fittings is shown in Table 4.11 and it was computed using the formula presented in Section 4.2.1. Figure 4.56(a) gives the time series of recent daily cases of COVID-19 in France which shows stationarity with rolling values (window = 12) appearing to be varying slightly. Also, Augmented Dickey-Fuller Test was used to test for stationarity

of the time series by accepting the alternative hypothesis with a p-value less than the 5% (p-value = 0.02) critical value, so it is 95 percent certain that this is a stationary series. Also, Figure 4.56(c) shows the plotting of three French departments (Nord, Paris and Essonne) with more prevalent COVID-19 hospitalization cases and Figure 4.56(d) shows the fitted curve of two of the French departments (Paris and Seine-Maritime) while all departments have RMSE in the interval $0.51 \leq RMSE \leq 17.38$ with Essonne department having the highest RMSE and Lozère department having the lowest RMSE. Other RMSE values are in Table 4.11. A deep learning forecasting result using GRU for France data between the beginning of the pandemic in France till September 3 2021 by training 80% of the data and testing 20% is presented Figure 4.56(b). The predicted cases curve values decline over the whole-time.

Table 4.11: RMSE confidence interval for all French departments for the fitness curve of the four functional data.

	RMSE before vaccination started	**RMSE after vaccination has started**
Hospitalized	$0.51 \leq RMSE \leq 17.38$	$1.00 \leq RMSE \leq 18.00$
ICU	$0.05 \leq RMSE \leq 2.60$	$0.35 \leq RMSE \leq 5.20$
Daily return home	$0.25 \leq RMSE \leq 12.49$	$1.10 \leq RMSE \leq 17.50$
Daily deceased	$0.04 \leq RMSE \leq 4.52$	$0.32 \leq RMSE \leq 4.10$

4.5.2 Data smoothing

The first step in analyzing functional data is to smooth the curves. In this section, different smoothing techniques which shall be illustrated and basic explanation of the techniques deployed for smoothing the functional data will be presented. The mean of the data set which corresponds to elastic mean and the cross-sectional mean is plotted (see Figure 4.57), which corresponds to the karcher-mean under the $L2$ distance.
The karcher-mean has an important role in the warping framework and is used to align the functions in order to improve the matching of features (peaks) across functions for a given set of warping functions $\partial_1, \partial_2, \cdots, \partial_n \in \Delta$ as

$$\widetilde{\partial}_n = argmin_{\partial \in L2=S\Delta} \sum_{i=1}^{n} dist(\partial, \partial_i)^2,$$

where S is a quotient space, Δ is the differential geometry, $\partial \in \Delta, \partial_i$ is the set of warping functions [246].
It was observed that the elastic mean better captures the geometry of the curves compared to the standard $L2$ mean for some of the functional data set considered, since it is not affected by the deformations of the curves. This phenomenon can be seen in Figure 4.57(a)–(c) and Figure 4.57(e)–(g), but Figure 4.57(d) and Figure 4.57(h) show a bad shape for elastic mean. The B-spline smoothing technique was used in smoothing the functional data where the number of basis function $n = 7$ which was discovered that it best suit for the modelling of the data as number of elements. The result of this smoothing technique can be seen in Figure 4.58.

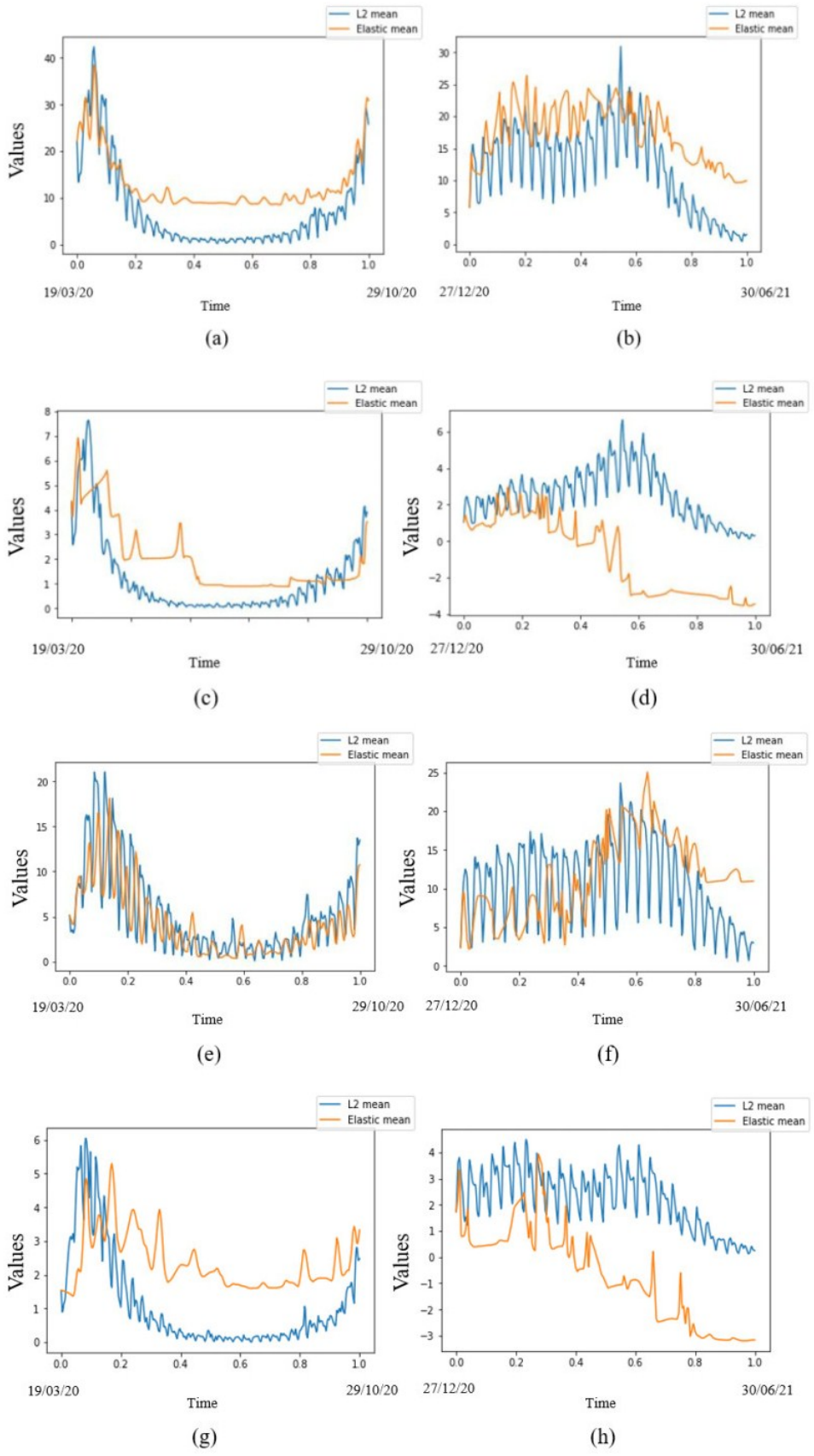


Figure 4.57: (a,c,e,g) - hospitalized, ICU, daily return home and deceased cases before start of vaccination, (b,d,f,h) - hospitalized, ICU, daily return home and deceased cases.

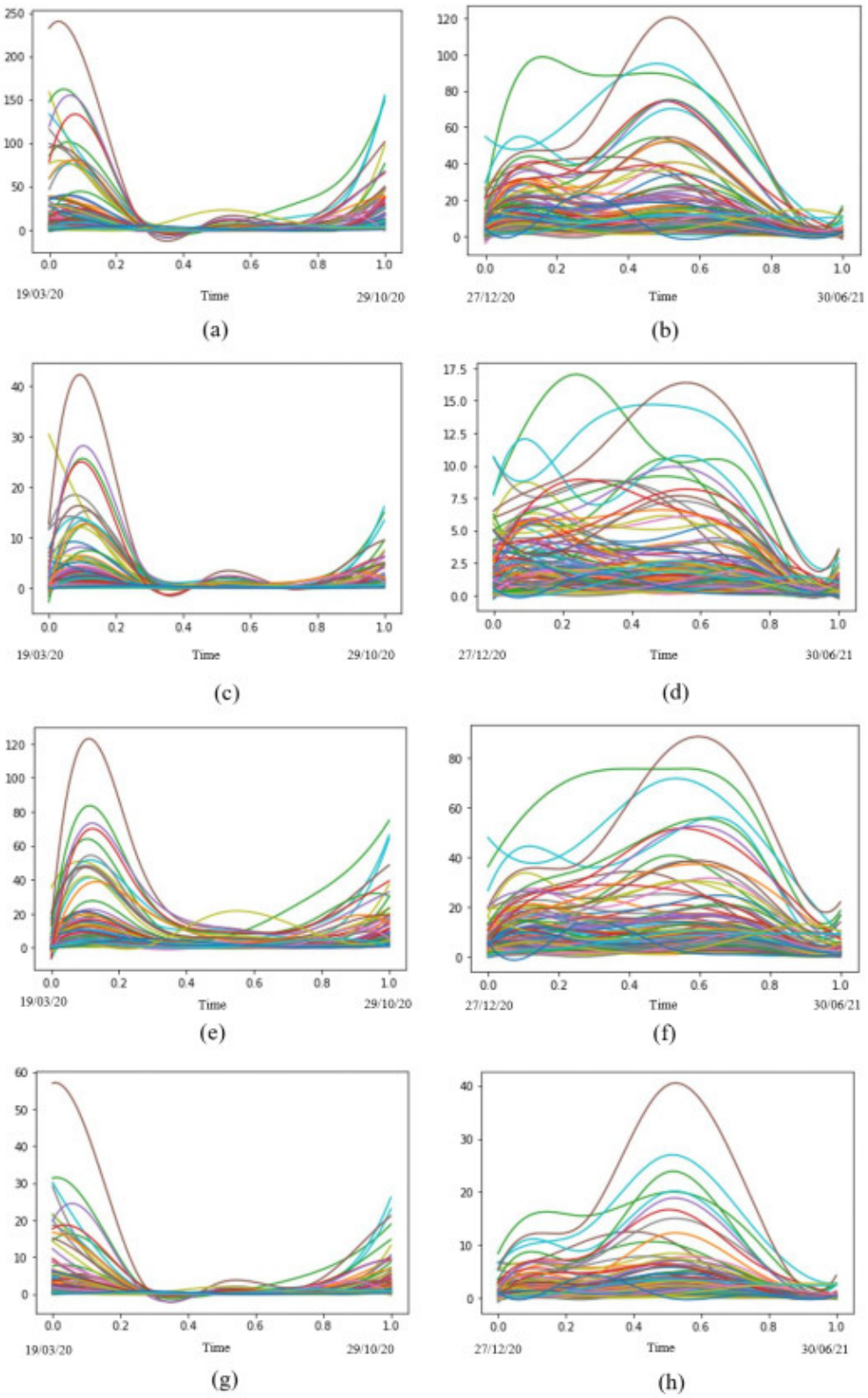


Figure 4.58: Smoothed curves for the shape of COVID-19 epidemic in all departments in France: (a) hospitalized cases, (b) hospitalized when vaccination has started, (c) daily deceased, (d) daily deceased when vaccination has started, (e) daily return home, (f) daily return home when vaccination has started, (g) ICU cases and (h) ICU cases when vaccination has started.

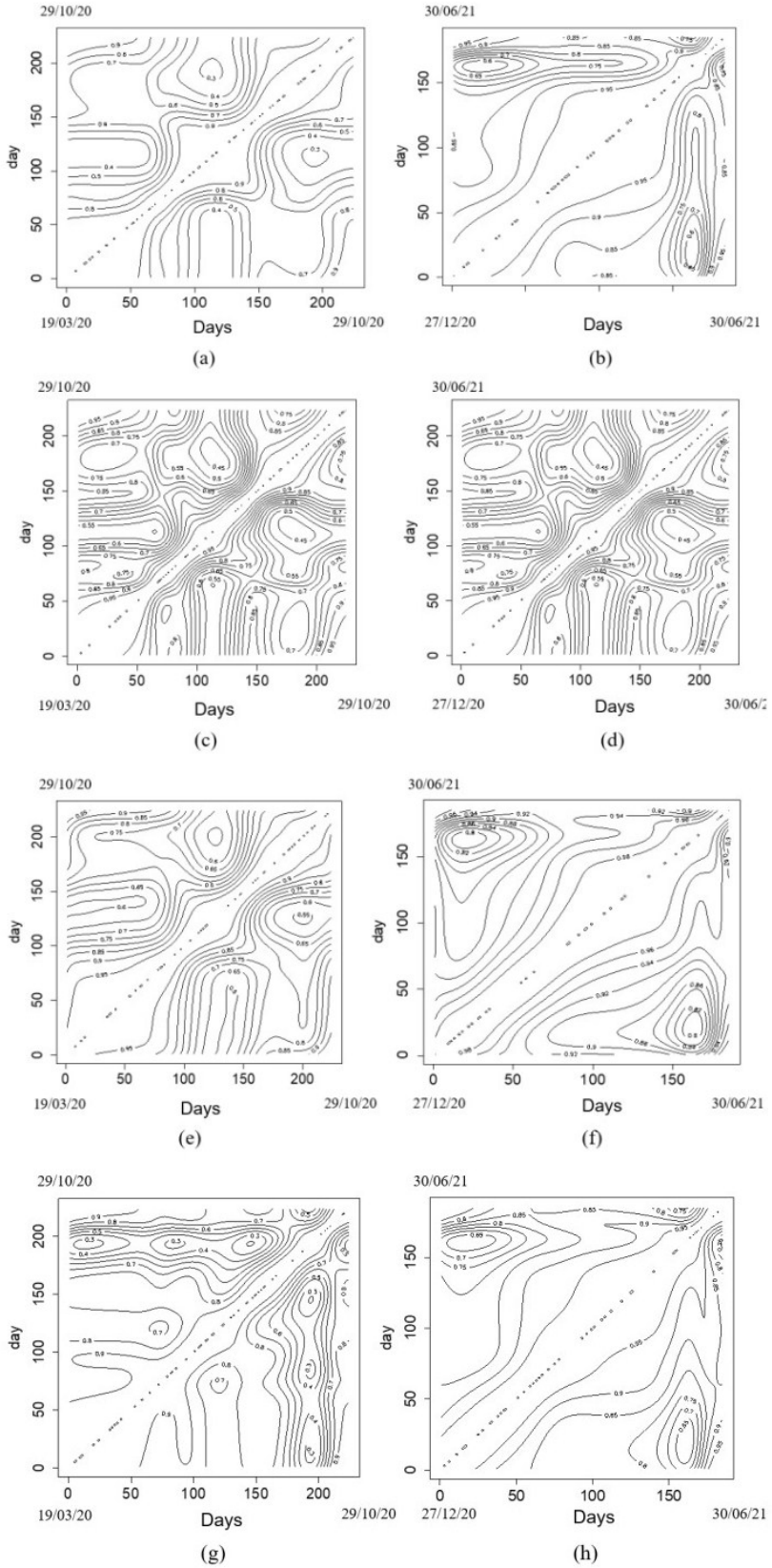


Figure 4.59: Correlation coefficients between all French departments contour plot. (a) hospitalized cases, (b) hospitalized when vaccination has started, (c) ICU cases, (d) ICU cases when vaccination has started, (e) daily return home, (f) daily return home when vaccination has started, (g) daily deceased and (h) daily deceased when vaccination has started.

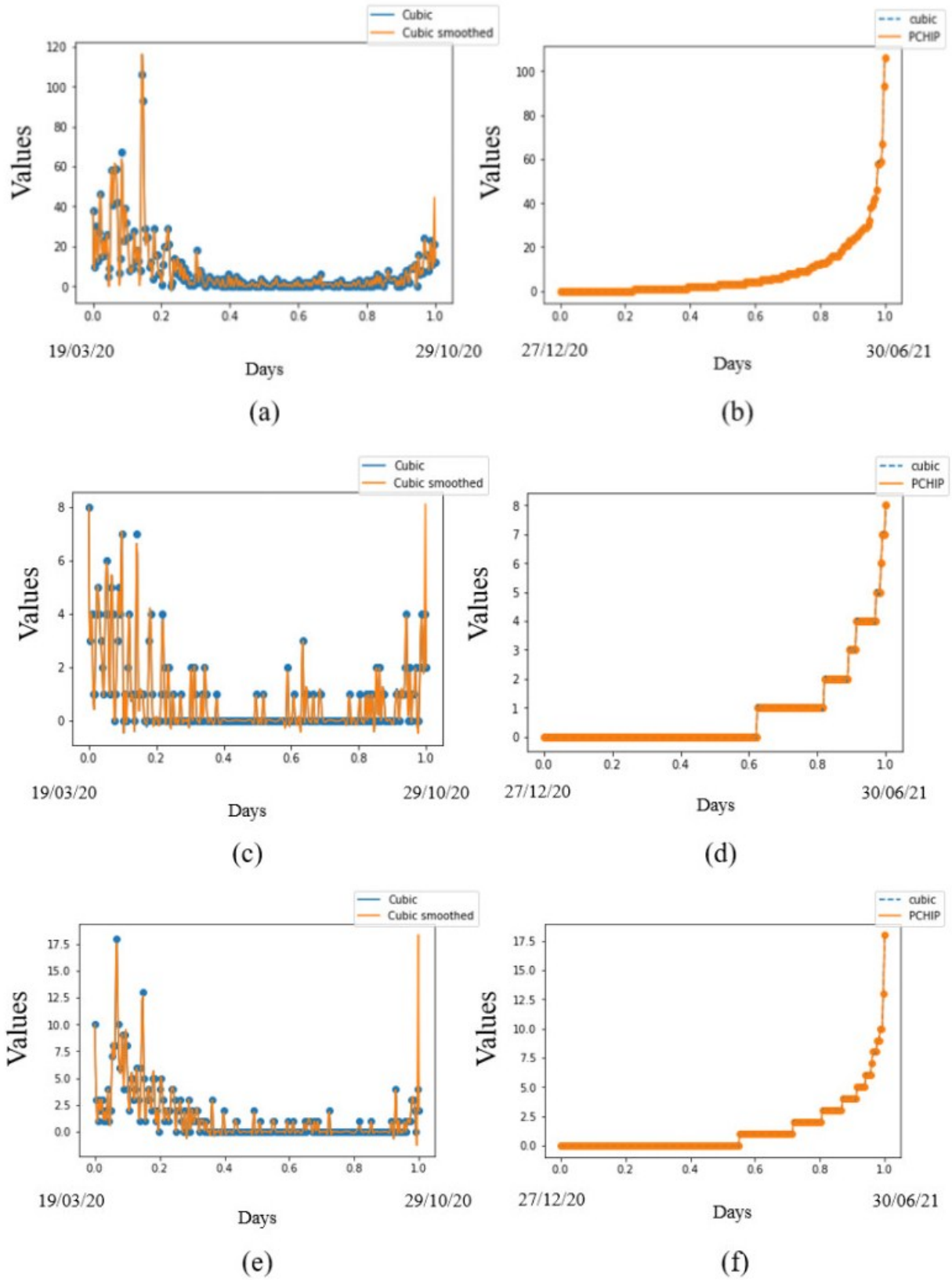


Figure 4.60: (a) Hospitalized cases interpolation smoothing, (b) hospitalized monotone smoothing, (c) ICU cases interpolation smoothing, (d) ICU cases monotone smoothing, (e) daily deceased interpolation smoothing and (f) daily deceased monotone smoothing.

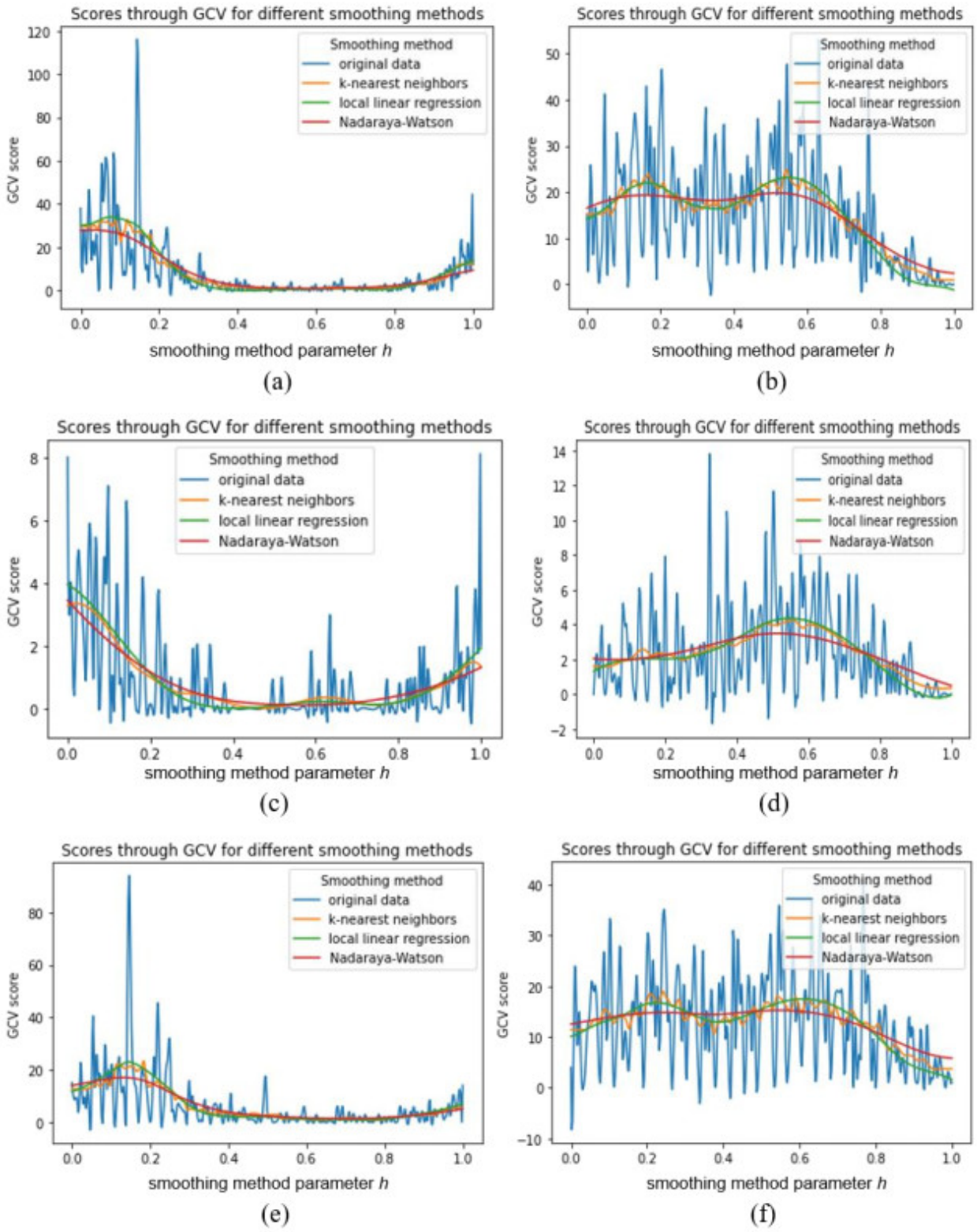


Figure 4.61: Kernel smoothing method for (a) hospitalized cases, (b) hospitalized when vaccination has started, (c) ICU cases, (d) ICU cases when vaccination has started, (e) daily return home and (f) daily return home when vaccination has started.

In Figure 4.59, the correlation coefficient between all the departments in France based on the

functional data in consideration is plotted, in order to see how well the data is well correlated between the departments and it was observed that there is a high correlation between various departments with except in few cases where low correlation is observed as we can see in the contour plots presented in Figure 4.59(a)–(h).
The further smooth the data, spline interpolation of order 3 and smoothing the interpolation using the smoothness parameter equal to 1.5 in the cubic spline smoothing was used. Also the monotone technique and a piece-wise cubic Hermite interpolating polynomial (PCHIP) was used. The visualization of the results is in Figure 4.60.
From graphs Figure 4.603(a)–(f) one can deduce that cubic spline smoothing curves exhibit oscillations and oscillations are important to know the low and high data thresholds in case the consecutive data points experience a significant change in slope. It was also observed that PCHIP is smooth and non-oscillatory despite some sharp increase as the U-shape of the curve deepens.
Figure 4.61 shows the Kernel smoothing for the smoothed curves of the functional data set for three different smoothing methods to see how cross validation score varies over a range of different parameters used in smoothing methods. The essence of this smoothing is to estimate the smoothing parameter h that better represents functional data. It has been selected by Generalised Cross-Validation criteria (GCV). The results show a comparable behavior of these scores by varying the smoothing parameter h.

4.5.3 Functional principal component analysis (FPCA)

Functional principal component analysis (FPCA) is a dimension reduction with high correlation in functional data analysis which completes the statistical tools used in the modelling of epidemiology data. In this section, the 4 PCs values plot is presented throughout the days considered and the principal component scores plot for all the different departments providing functional data being before vaccination started and during vaccination.
In Figure 4.62(a) it was observed that PC 1 peaked in the early days of the pandemic between February and March 2020 and then there was a decline after about 50 days becoming stationary till day 150 possibly due to mitigation measures promulgated during this period. The same phenomenon has been observed for PC 2. In Figure 4.62(a), PC 4 shows a sinusoidal shape, peaked at day 100 which is around June 2020 with least values at day 30 and day 180 which are respectively in March and August 2020. Figure 4.62(b) shows the same sinusoidal shape for PC 4 and same shape for PC 3 but with a drift in the observation with a difference between the dynamics of hospitalization cases before and after vaccination has started in France. PC 1 in Figure 4.62(b) shows a decline across the infective period which may be due to the aggressive vaccination campaign in the country.
In Figure 4.62(c) it was observed that from day 50 (around April 2020) till day 150 (around July 2020), the PC 1 value which is the major PC is stable throughout this period of various confinement measures in France and all PCs tend to show increasing behavior after the confinement measures have been relaxed and in Figure 4.62(d), PC 1 has strictly positive values while PCs 2–4 show negative values between February to June 2021. In Figure 4.62(e) PC 1 peaked with a positive value at the beginning of the pandemic in France which validates the percentage of recovery in France (about 71% as at 14/11/2021) while PC 1 in Figure 4.62(f) shows a positive decline across the days considered, with a disparity between the period of vaccination and without vaccination.
On the y-axis of Figure 4.62(g) and Figure 4.62(h), it was observed that this is the only result

with low values for the PCs because the deaths due to COVID-19 in France remain at a low level (about 2% as at 14/11/2021), while all PCs show almost the same pattern as what was observed previously for the other variables.
In Table 4.12, the PCs variance proportion is presented and it was observed that PC 1 is the most important principal component.

Table 4.12: PCA variance proportion for 4 PCs before and after vaccination started in France.

	Before				**After**			
	PC1	**PC2**	**PC3**	**PC4**	**PC1**	**PC2**	**PC3**	**PC4**
Hospitalized	0.945	0.039	0.008	0.005	0.938	0.041	0.012	0.004
ICU	0.960	0.028	0.009	0.001	0.962	0.023	0.008	0.004
Daily return home	0.925	0.045	0.015	0.007	0.953	0.030	0.009	0.004
Daily deaths	0.965	0.017	0.013	0.003	0.914	0.055	0.016	0.010

For functional principal component scores and clusters, the focus will be on the departments where the pandemic is most prevalent, as well as on PCs 1 and 2, while ignoring the other PCs. Recall that French departments have code numbers, for example, Nord is code number 59, and these codes will be used this to described the departments as well as in the visualization of results shown in Figures 6 and 7. For a complete list of all French department code numbers see [96]. In Figure 4.63a, the Paris department (code number 75) and Nord department (code number 59) have a positive score in PC 1 and negative score in PC 2 while the Essonne department is positive in both PCs. In Figure 4.63b, the Paris department and Essonne department (code number 91) are negative in both PCs while the Nord department is positive in PC 2 with the highest score and negative in PC 1. Figure 4.63c, Nord and Essonne departments are negative in PC 2 but positive in PC 1 while the Paris department is positive in both PCs.
The Paris department and Essonne department are negative in both PCs in Figure 4.63d while the Nord department is positive in PC 2 and negative in PC 1. In Figure 4.63e, Paris and Nord departments have positive scores in both PCs while the Essonne department is negative in PC 2 and positive in PC 1. Nord department has the highest positive score in PC 1 for Figure 4.63f and negative for PC 1, Paris department is positive in PC 2 and negative in PC while Essonne department is negative in both PCs. The Paris department has the highest positive score in PC 1 and negative in PC 2 in Figure 4.63g, Nord department is positive in both PCs while Essonne department is negative in PC 2 but positive in PC 1. Finally, in Figure 4.63h while Nord department is positive and highest in PC 2, Paris department is the lowest with negative score in PC 2. Both departments are negative in PC 1. Essonne department is positive in PC 2, but negative in PC 1.
This description shows that there is a difference between the vaccination period in France and the period when measures like lockdown, social distancing etc. were only used to control the spread of the virus despite the fact that it has been proven medically that people can be vaccinated and still be infected. The diagrams in Figure 4.63a–h show the same shift toward positive PC 1 values. These shifts demonstrate the effect of various mitigation measures differences in the departments and based on population and migration in this area, particularly the five departments (Mayotte, Guadeloupe, La Réunion, Guiana and Martinique) outside France's metropole where the rules in France are not strictly enforced. It also demonstrates that PC 1 is the most important PC, from which the majority of the analysis information can be obtained.

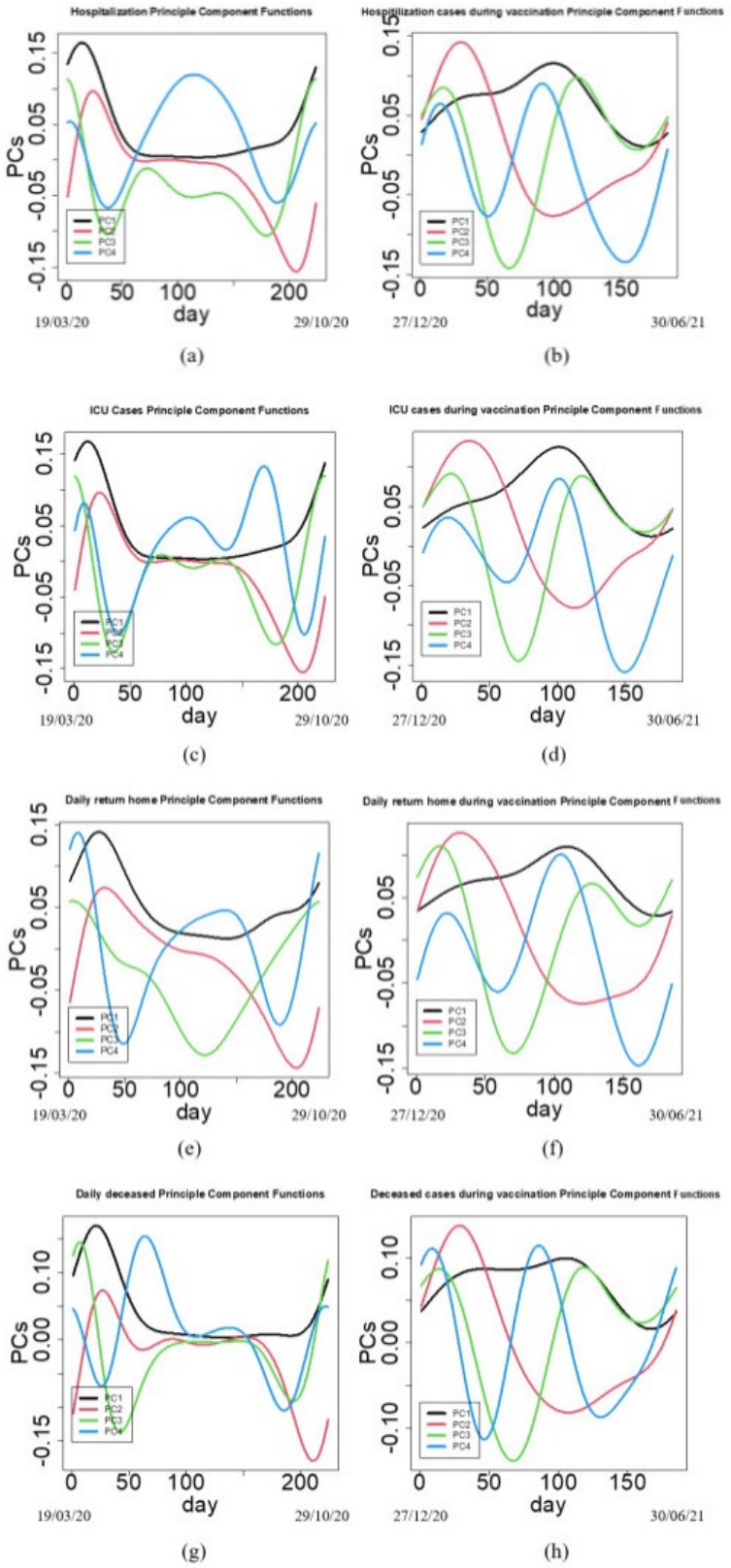


Figure 4.62: Functional PCs for different functional data before the start of vaccination (19/03/2020–29/10/2020) and when vaccination has started (27/12/2020 to 30/06/2021): (a) hospitalized cases, (b) hospitalized when vaccination has started, (c) ICU cases, (d) ICU cases when vaccination has started, (e) daily return home, (f) daily return home when vaccination has started, (g) daily deceased and (h) daily deceased when vaccination has started.

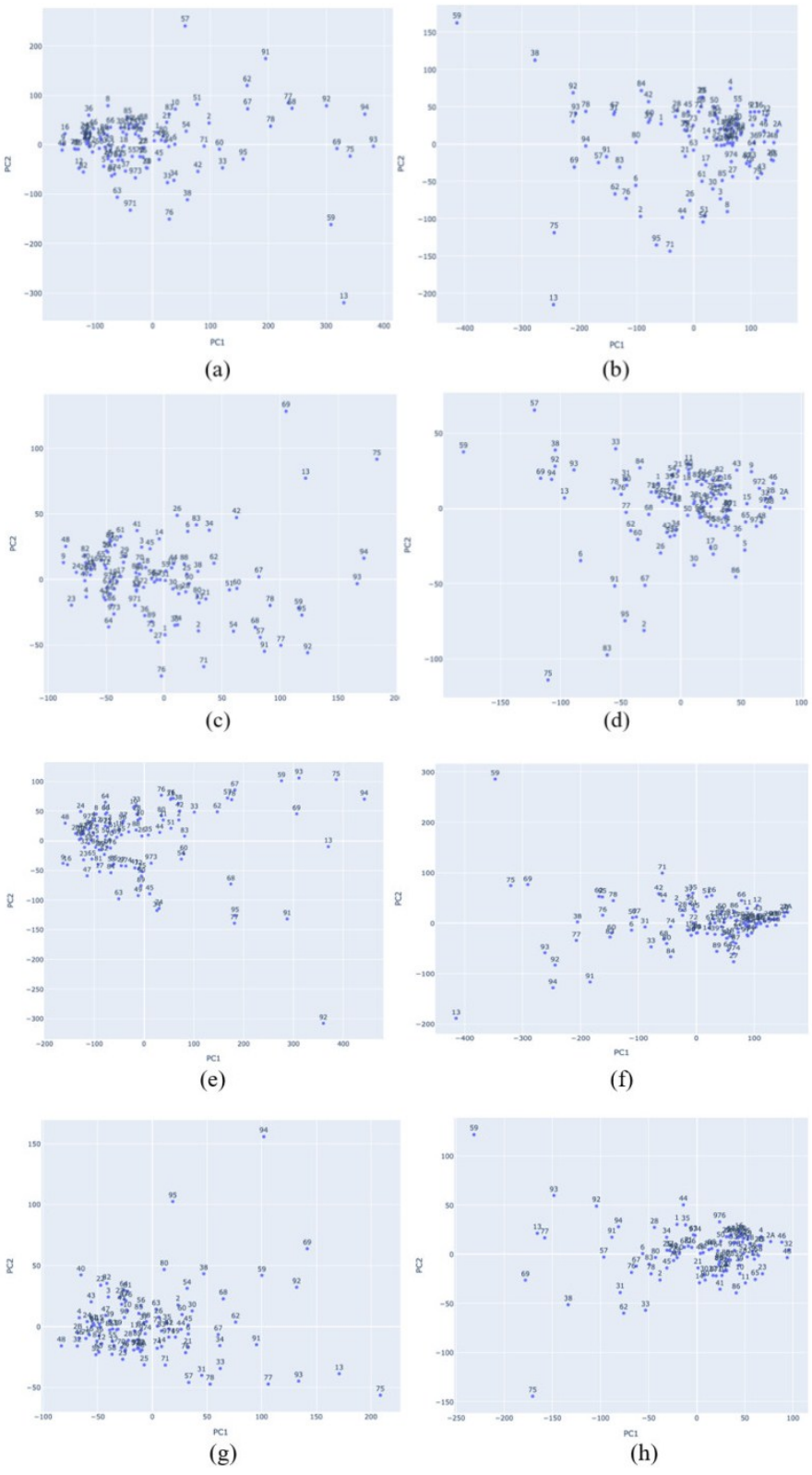


Figure 4.63: FPCA scores for different functional data before the start of vaccination (19/03/2020–29/10/2020) and when vaccination has started (27/12/2020–30/06/2021). (a) hospitalized cases, (b) hospitalized when vaccination has started, (c) ICU cases, (d) ICU cases when vaccination has started, (e) daily return home, (f) daily return home when vaccination has started, (g) daily deceased and (h) daily deceased when vaccination has started. Note that the numbering of points on the diagram are codes for each French department.

4.5.4 Canonical correlation analysis (CCA)

Canonical correlation is a statistical analysis method that is used to simultaneously correlate several metric dependent variables and several metric independent variables measured on or observed with similar experimental units. It helps to find linear combinations within a data set with the goal of maximizing the correlation between these linear combinations. The variables used in this section are presented in Table 4.13. X are the variables listed in the first row of Table 4.13 i.e., total number of hospitalizations, daily return home, deceased and ICU cases for all departments before and after vaccination has started in France and Y variables are the response variables described earlier as presented in the first column of Table 4.13. The visualization of the results can be found in Figure 4.64 and Figure 4.66 and also the correlation scores is presented in tabular form (see Table 4.13).
The helio plot in Figure 4.64 depicts the relationships between the different departments in France, as well as the epidemiology variables and control measures (vaccination).
Figure 4.64(a), Figure 4.64(c) and Figure 4.64(e) showed a negative correlation between the epidemiology variables and the hospitalization, ICU, and daily return home cases across departments, whereas Figure 4.64(g) showed a positive correlation between the epidemiology variables and the deceased cases across departments. Figure 4.64(b), Figure 4.64(d) and Figure 4.64(f) confirm the effect of vaccination on the number of hospitalizations, ICU, and daily return home during vaccination, demonstrating a positive correlation and the effect of this control measure, whereas Figure 4.64(h) shows a negative correlation, indicating a negative relationship between the deceased and the vaccination introduced. The results presented in Figure 4.66 show the linear relations in the scatter plot as most of the variables show 95% significance level and from Table 4.13 there is high correlations between the variables considered. The Figure 4.65 presents the redundancy between the canonical variates.
The canonical correlation is tested and the null hypothesis is when the canonical correlation is equal to zero. Figure 4.66(a) shows hospitalized cases with p-value < 0.05 for all canonical variate, Figure 4.66(b) shows hospitalized when vaccination has started with p-value < 0.05 except the last canonical variate with p-value $= 0.88$, Figure 4.66(c) shows ICU cases with p-value < 0.05 for all canonical variate, Figure 4.66(d) shows ICU cases when vaccination has started with p-value < 0.05 except the last two Canonical variate with p-value $= 0.68$ and p-value $= 0.87$ respectively, Figure 4.66(e) shows daily return home with p-value < 0.05 for all canonical variate, Figure 4.66(f) shows daily return home when vaccination has started with p-value < 0.05 except the last two canonical variate with p-value $= 0.14$ and p-value $= 0.34$ respectively, Figure 4.66(g) shows daily deceased with p-value < 0.05 except the last canonical variate with p-value $= 0.08$ and Figure 4.66(h) shows daily deceased when vaccination has started with p-value < 0.05 except the last two canonical variate with p-value $= 0.08$ and p-value $= 0.46$ respectively.

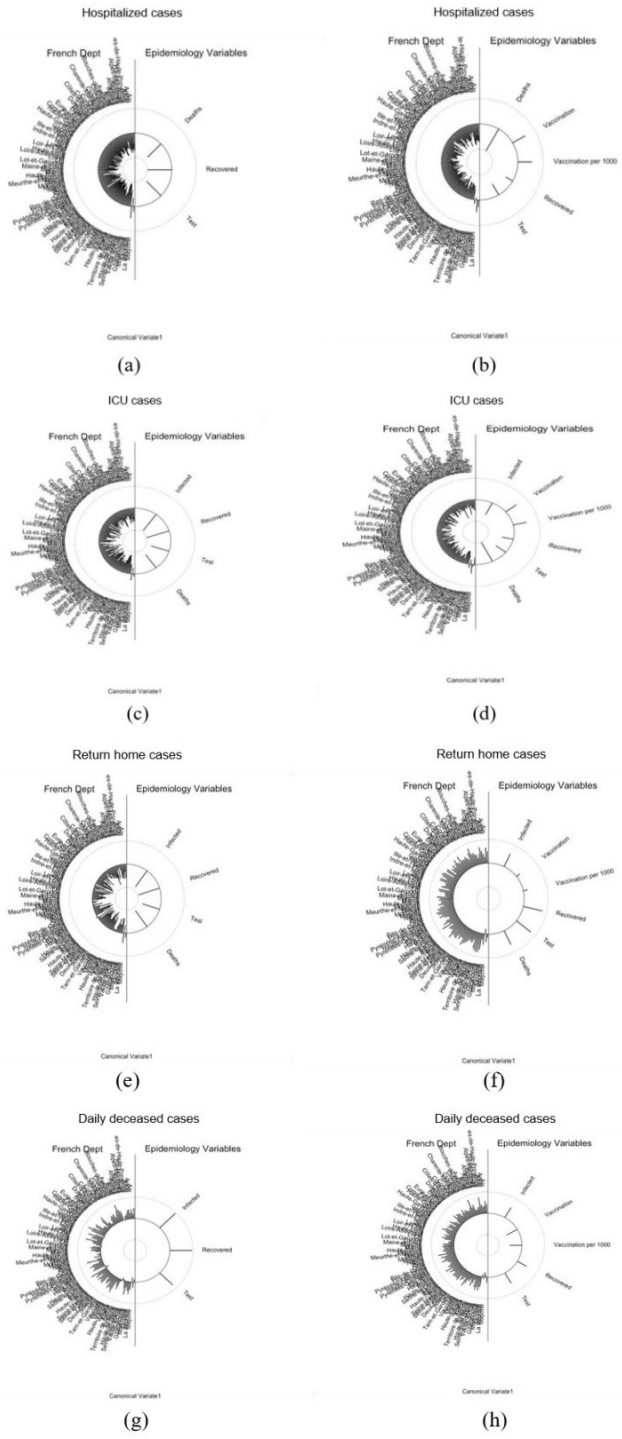


Figure 4.64: Helio plot for the correlation of French departments for (a) hospitalized cases, (b) hospitalized when vaccination has started, (c) ICU cases, (d) ICU cases when vaccination started, (e) daily return home, (f) daily return home when vaccination started, (g) daily deceased and (h) daily deceased when vaccination started.

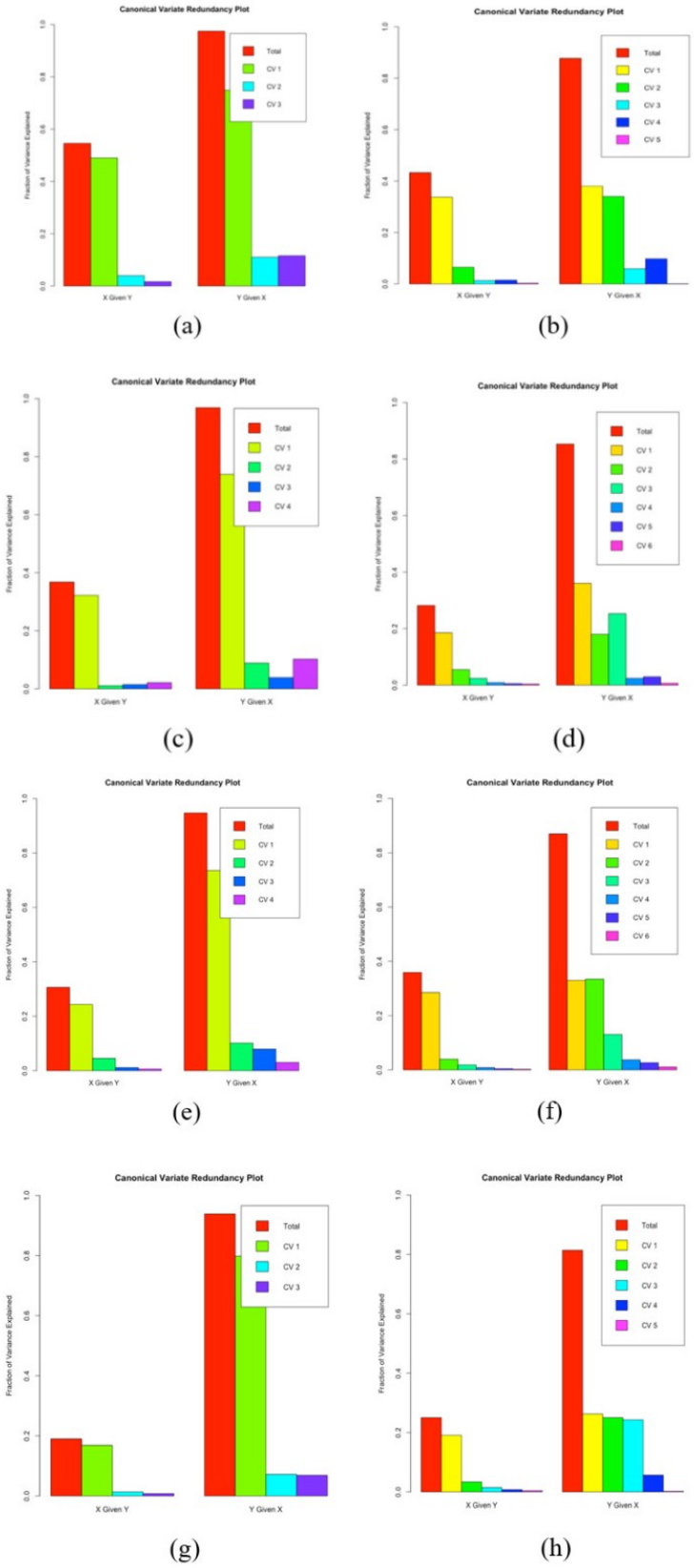


Figure 4.65: Canonical variate redundancy plot for (a) hospitalized cases, (b) hospitalized when vaccination has started, (c) ICU cases, (d) ICU cases when vaccination has started, (e) daily return home, (f) daily return home when vaccination has started, (g) daily deceased and (h) daily deceased when vaccination has started.

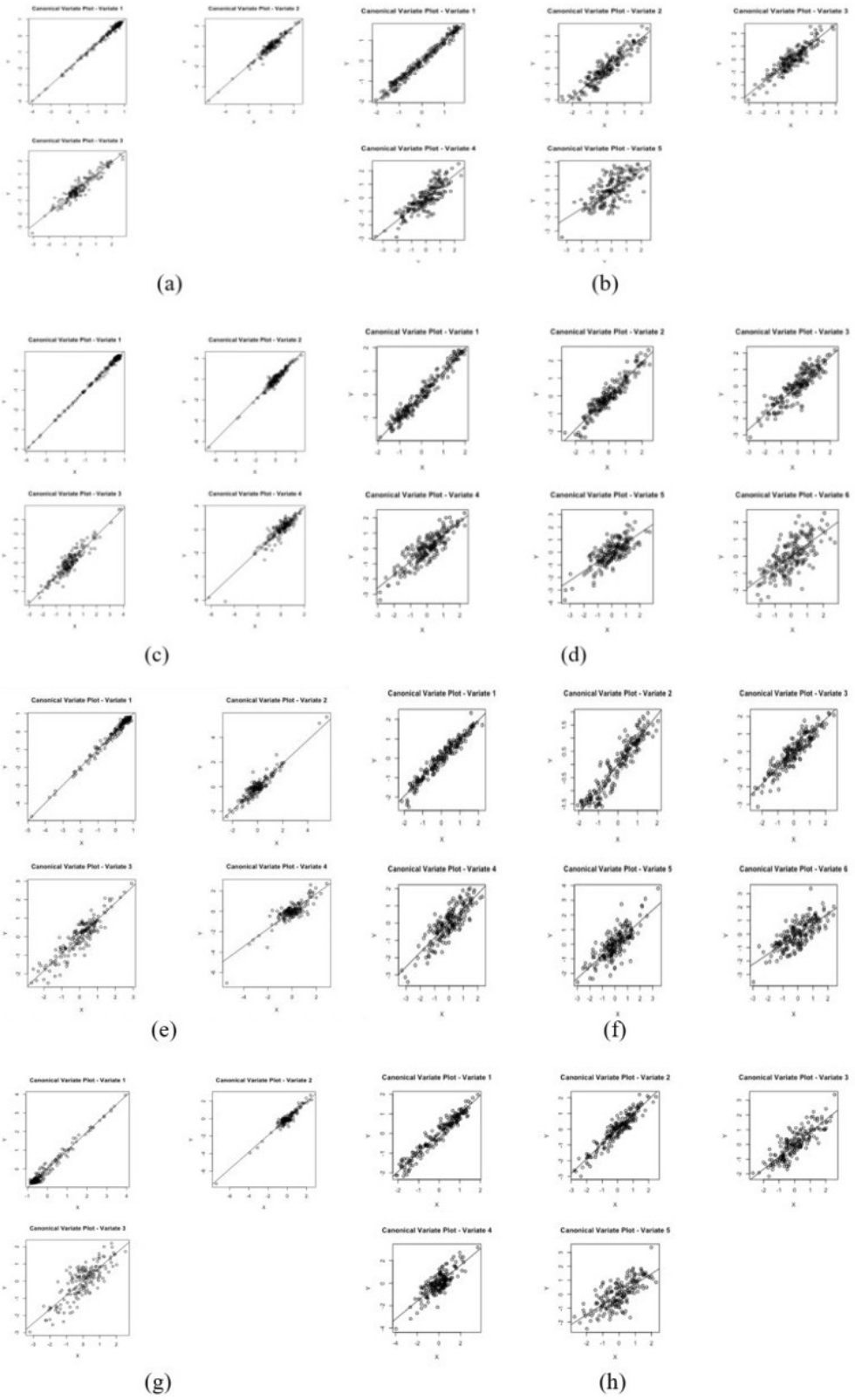


Figure 4.66: Canonical correlation visualization of (a) hospitalized cases, (b) hospitalized when vaccination has started, (c) ICU cases, (d) ICU cases when vaccination has started, (e) daily return home, (f) daily return home when vaccination has started, (g) daily deceased and (h) daily deceased when vaccination has started.

Table 4.13: Canonical correlation scores for the variables before and after vaccination started in France.

	Before				After			
	Hospitalized	**ICU**	**DRH**	**Daily deaths**	**Hospitalized**	**ICU**	**DRH**	**Daily deaths**
Deaths	0.996	0.926	0.859	-	0.989	0.689	0.745	-
Recovered	0.970	0.973	0.941	0.961	0.838	0.865	0.846	0.797
Test	0.950	0.937	0.911	0.816	0.685	0.750	0.776	0.736
Vaccination	-	-	-	-	0.924	0.942	0.939	0.936
Infected	-	0.998	0.992	0.987	-	0.980	0.971	0.970
Vaccination/1000	-	-	-	-	0.901	0.885	0.917	0.841

4.5.5 Clustering method

The clustering of functional data is one method that statisticians are always interested in and in this section the K-means and fuzzy K-means techniques whose algorithm is already in Python was used. These methods will enable us to visualize how various departments are clustered based on our functional data and to give it the best interpretation based on their geographical location. The basic function used for the K-means clustering is a B-spline and results of the clusters are presented. The result in the cluster form and also on the map of France with indication of the membership to the 3 (i.e., K=3) clusters (0, 1 or 2) to get a clearer view of the result are presented. Even if the value of K is increased, it won't improve the results as K is optimal. Only the result for two cases (daily hospitalized and daily deceased) for the period before vaccination begins in France and two cases (daily return and ICU cases) for the period when vaccination has started in France is presented. Figure 4.67(a)–(d) presents the clusters (0, 1 or 2) that each French department belongs to. The result clustered French departments outside metropole or the French hexagon to the same clusters which of course are not binded with mitigation measures and rules used in departments within France [96]. Also, departments close to Paris are in the same cluster which is the same with departments having the same trend of the pandemic prevalence as presented in Figure 4.67(a)–(d). In Figure 4.67(e) and Figure 4.67(g) the same pattern for the data points in the clustering were observed which means that the same way that there are more hospitalization cases before vaccination, also there are more people returning home during vaccination which affirm the fact that the vaccination campaign in France has helped to mitigate the spread of the disease.
Figure 4.67(f) and Figure 4.67(h) is a bit tricky because of the pattern they followed but it is not surprising that there is more deceased in most departments before vaccination period and less ICU cases during the vaccination period. These patterns of having more cluster points attaining the highest peak in PC 1 is distinct in the results presented.

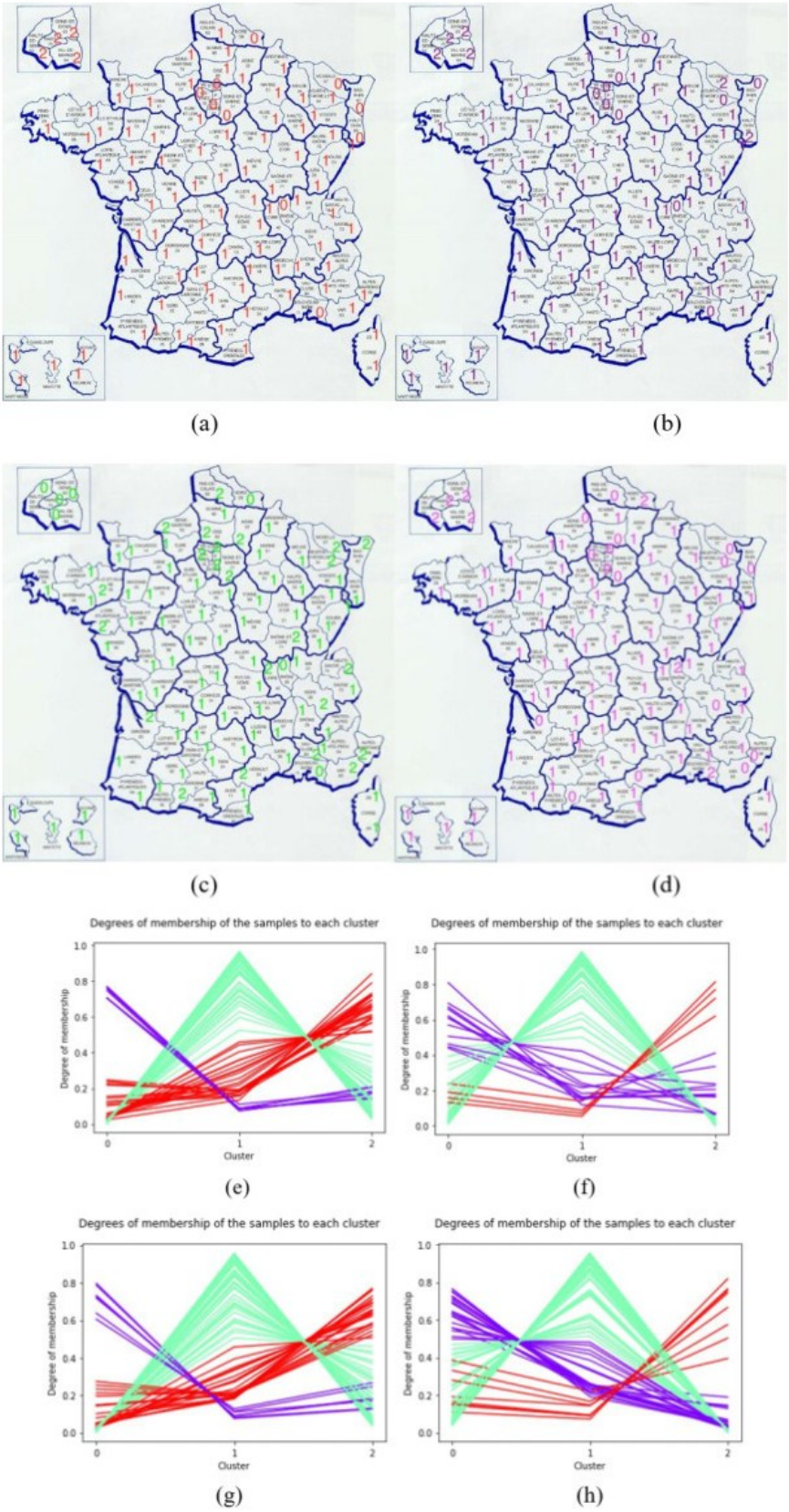


Figure 4.67: Clustering of all departments in France using K-means: (a) hospitalized, (b) daily deceased, (c) daily return home when vaccination has started and (d) ICU cases when vaccination has started, and fuzzy K-means, (e) clusters for hospitalized, (f) clusters for daily deceased, (g) clusters for daily return home when vaccination has started and (h) clusters for ICU when vaccination has started.

4.5.6 Function to scalar linear regression modeling

This section describes the used function to scalar linear regression model to predict two of our response variables which is of the form:

$$Y_{\text{Deaths}} = \int_1^{177} \theta(t) X_{\text{fdDataHospitilization}}(t)dt + \varepsilon$$

$$Y_{\text{Tests}} = \int_1^{177} \theta(t) X_{\text{fdDataHospitilization}}(t)dt + \varepsilon$$

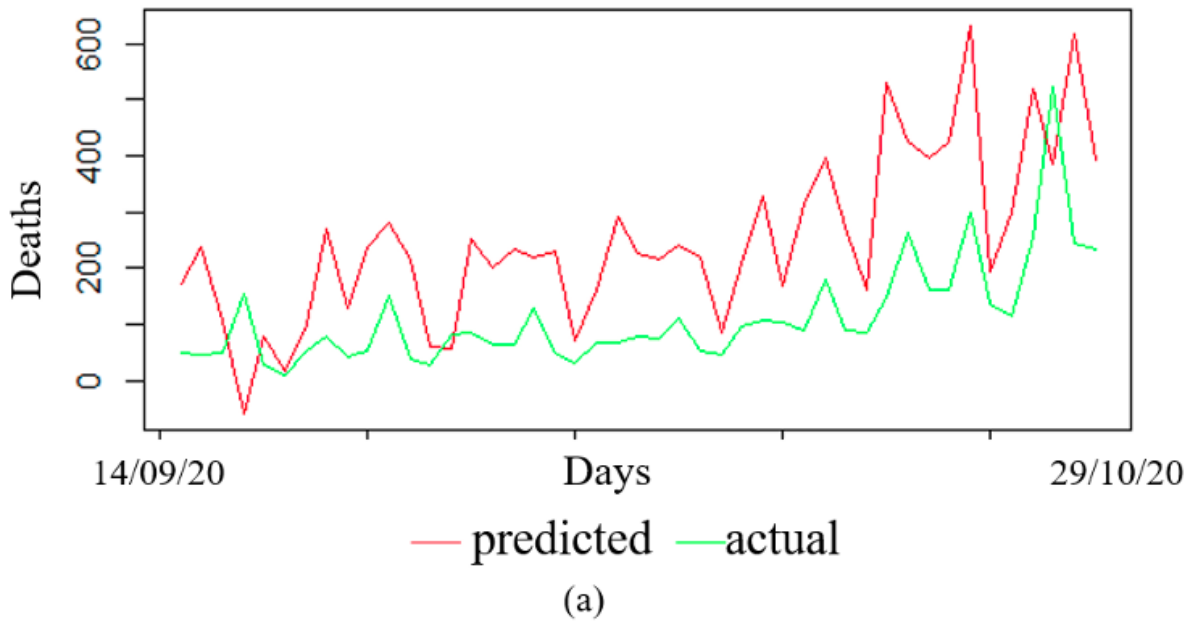
The data considered in this section are data before vaccination started in France and 80% was trained and 20% was tested. By applying the model presented Equation (2.23), in this case, the functional data is X (hospitalization) which is a curve whose prediction is linked to a scalar Y (number of deaths and tests between 19/03/20 – 13/09/20 which is the time interval representing $[1, 177]$) response variable. The time interval for in which the response was predicted is from 14/09/20 – 29/10/20 which represent $[178, 225]$ (the test data). The visualization of the results is presented in Figure 4.68 and the tabular form of the numerical results can be found in Table B.3 in the Appendix B. The prediction affirms the fact that the relaxation in the mitigation measures during the studied period increases the number of deaths and tests in France, the predicted results of deaths being systematically higher than the observed values as seen in Figure 4.68a and Table B.3 in the Appendix B.

4.5.7 Function-on-function linear regression modeling

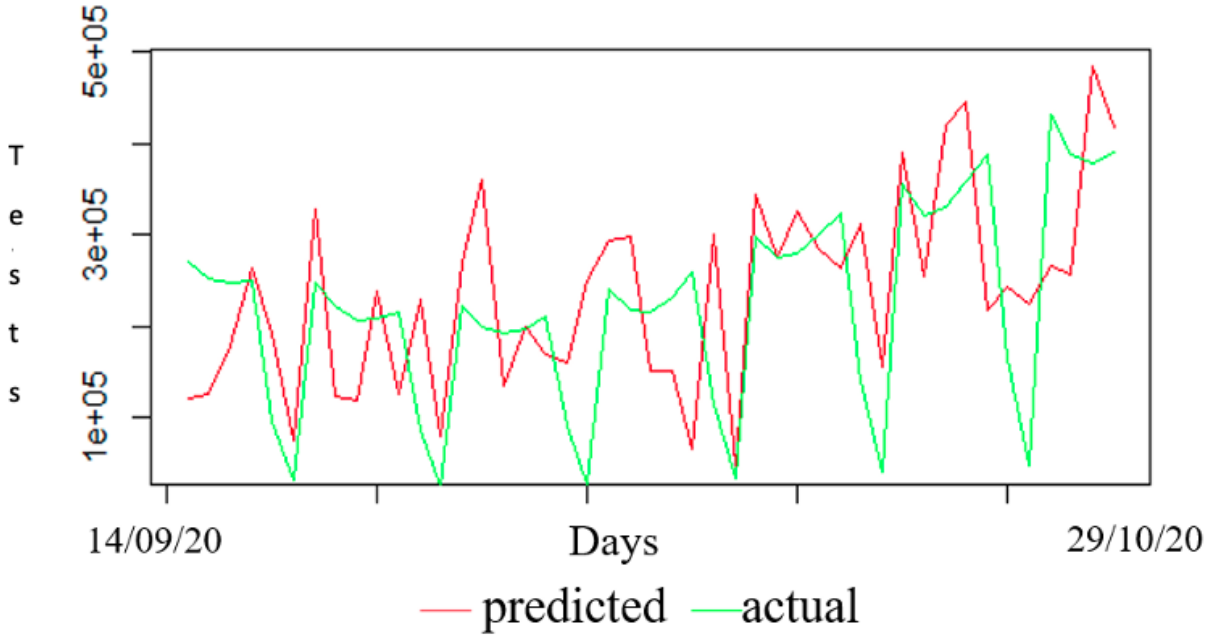
The modeling of the functional data using function-on-function linear regression of the form of Equation (2.25) i.e.,

$$y_{\text{AllfdData}}(t) = \beta_0(t) + \int_1^{101} \beta(s,t) x_{\text{AllfdData}}(s)ds + \varepsilon$$

by using the functional data curves of 101 days (predictors) to predict another set of curves of 101 days (response) while also estimating the slope $\beta(s,t)$, whose results in considered cases are presented in 3D diagrams of Figure 4.69.
Figure 4.69(a) shows hospitalized cases with the slope on the interval $-2.799063 \leq \beta(s,t) \leq 1.980147$, Figure 4.69(b) shows hospitalized when vaccination has started with the slope on the interval $-1.501887 \leq \beta(s,t) \leq 1.076421$, Figure 4.69(c) shows ICU cases with the slope on the interval $-1.0733846 \leq \beta(s,t) \leq 0.841100$, Figure 4.69(d) shows ICU cases when vaccination has started with the slope on the interval $-0.5646148 \leq \beta(s,t) \leq 0.3661280$, Figure 4.69(e) shows daily return home with the slope on the interval $-0.6755000 \leq \beta(s,t) \leq 0.7030529$, Figure 4.69(f) shows daily return home when vaccination has started with the slope on the interval $-0.4333295 \leq \beta(s,t) \leq 0.4300995$, Figure 4.69(g) shows daily deceased with the slope on the interval $-0.3277864 \leq \beta(s,t) \leq 0.4002531$ and Figure 4.69(h) shows daily deceased when vaccination has started with the slope on the interval $-0.3284866 \leq \beta(s,t) \leq 0.3641679$. The observation from all the figures in this section is that the 3D surfaces yield results whose shapes look roughly similar to the slope curve, functional predictors curve and functional response curve.



(a)



(b)

Figure 4.68: (a) Functional linear regression model prediction for number of deaths in France as response variable before vaccination begins and (b) Functional linear regression model prediction for number of tests in France as response variable before vaccination begins.

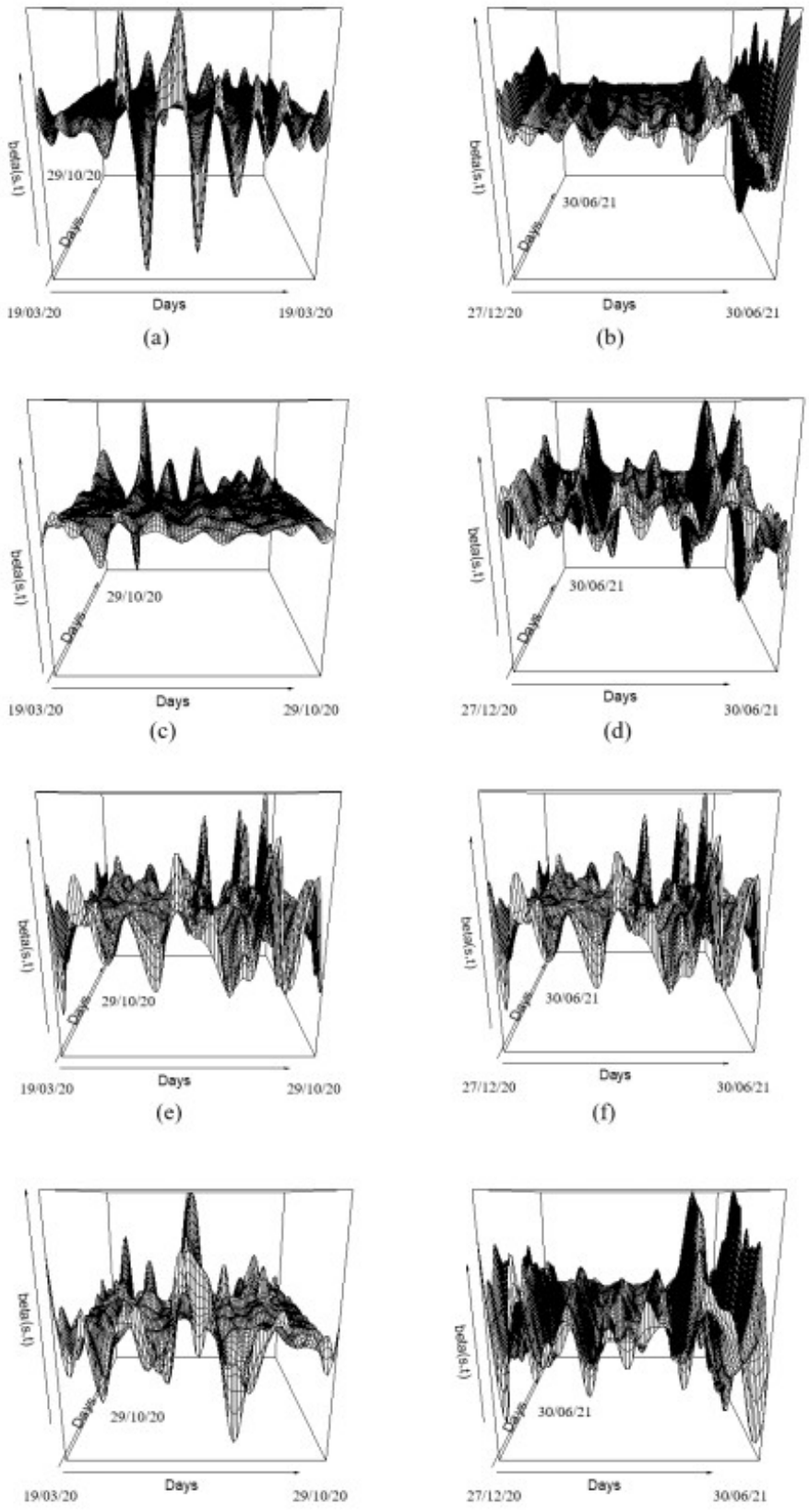


Figure 4.69: The 3D visualization of function-on-function regression for (a) hospitalized cases, (b) hospitalized when vaccination has started, (c) ICU cases, (d) ICU cases when vaccination has started, (e) daily return home, (f) daily return home when vaccination has started, (g) daily deceased and (h) daily deceased when vaccination has started.

4.6 Conclusion

Modern statistics applied to large-scale data analysis is computationally intensive, because many routine statistical tasks, such as distance matrix calculations, hierarchical clustering, principal component analysis, and so on, are difficult to calculate. Data analysis is widely used to gain new insights into viral disease evolution and to develop better mitigation methods and approaches. The current study demonstrated the ability of machine learning models and deep learning to accurately identify major components of viral disease such as COVID-19 and several determinants. Through the feature selection process, some of the regression models predicted some epidemiology variables and socio-economic variables, as dominant determinants in statistical modeling of the COVID-19 outbreak. The role of social inequalities in the current COVID-19 pandemic is becoming increasingly clear, both in terms of disease incidence and severity. Although the exact magnitude of these health disparities is difficult to quantify, the populations most at risk or vulnerable are frequently underrepresented in epidemiological studies and data. Many countries are in the grip of a severe and destructive recession. A significant number of firms and businesses have been downsized or closed.

We have proposed a set of different regression methods in order to find the best ones in the Kuwait context, both for daily new cases and deaths, without a priori about the degree of non-linearity and the stochastic structure of noise behind the data. Surprisingly, we discovered that often the best regression method was the support vector one and that the stochasticity of the data at the start of the two waves was the same. This result confirmed our choice of comparing several regression methods (exponential regression being the most commonly, and often the only one, chosen) and showed that once the trend due to epidemic dynamics and its seasonality has been removed, the random factors explaining the variations compared to the deterministic model of Bernoulli-Ross-McKendrick (namely the uncertainties related to the counting of new cases and deaths) were expressed through a similar noise for the first and the second wave. More, these two phases show a certain homogeneity in their stochastic structure at their beginning, because they have the same level for the variation coefficient (around 0.5), both for daily new cases number and for death number, and for the auto-correlation initial slope (-0.031 for the first phase and -0.038 for the second).

We have shown in this chapter that there exist correlations between the growth parameters directly linked to the occurrence of new cases of COVID-19 and socio-economic variables, in particular the ratio between the current health expenditures and gross domestic product percentage (CHE/GDP) anti-correlated with the basic reproduction time R_0, which shows the effectiveness of public health mitigation measures, even if they involve significant medico-economic costs. Larger perspectives are offered by combining this study with others on geoclimatic and demographic severity factors of the COVID-19 outbreak [102, 247] with the present socio-economic determinants, in order to obtain the most comprehensive and accurate picture of non-biological exogenous influences on the expanding COVID-19 pandemic.

In general, we have also been able to access the interplay between socio-demo-economic variables and epidemiologic variables for explaining how inequalities in these socio-demo-economic variables across countries, i.e., in terms of developed countries and developing countries responded to the COVID-19 outbreak. Also, from the results presented we can see that, there is a nexus between economic decisions taken by these countries and COVID-19 pandemic effects on the general economic activities in those countries.

Chapter 5

General Conclusion and Future Work

5.1 Scientific Contribution of the Research

The contributions made by this thesis are as follows:

- We have been able to provide the equation of the discrete dynamics of the epidemic's growth and obtain an estimation of the daily reproduction rates by using a deconvolution technique on a series of new COVID-19 cases from many countries. We established that the daily reproduction rates has four unique shapes but the U-shape is the most frequent evolution of daily reproduction rates, which confirms the comparison with the behavior of seasonal influenza.

- We have proved that socio-demo-economic determinants play a crucial role and that there is a relationship between these determinants during the exponential phase of the COVID-19 pandemic modeling.

- We have employed functional data analysis method to analyze the shapes of COVID-19 outbreak incidence rate curves in different departments in France and statistically group them into distinct clusters according to their shapes. The result reveals that pandemic curves often differ substantially across departments of same country, and we show that it is possible to analyze epidemic data obtained from a large sample concerning the incidence, mortality and exogenous or endogenous factors associated with the COVID-19 epidemic.

- We used the spectral analysis method to convert time (days) to frequency in order to analyze the peaks of frequency and periodicity of the time series data. Some of the peaks observed in the time series data correspond with the proven pattern of weekly peaks that is unique to the COVID-19 time series data and is due to the weekly incidence reporting.

- We proposed a new mathematical age-dependent SIGR (Susceptible–Infectious–Goneanewsusceptible–Recovered) model for the COVID-19 outbreak. We were able to simulate the differential dynamic behavior of the growth of infectious and immune populations, young and old in different countries, in order, for policy makers, to be able to have an idea of when to adjust the vaccine policy according to the age groups.

- We have been able to use different deep learning methods to retro-predict COVID-19 time series data from different countries. The results are closed and coincide often with the observed values of the time series based on our methods.

- We have been able to propose a continuous formulation of a unique maximum reproduction number estimate with an assumption that the epidemic curve is in form of the Gaussian curve and then compare the model with the discrete forms and the basic reproduction numbers observed during the contagiousness period considered. We also estimated the transmission rate from identification of the first inflection point of a wave on the curve of daily new infectious cases using the Bernoulli S–I (Susceptible–Infected) equation. We applied this new method to the real data from Cameroon COVID-19 outbreak both at national and regional levels and to the second wave COVID-19 outbreak for the world data which is a period the phenomena we are considering were observed.

- We developed an epidemic ageing model in human populations in order to estimate the lifespan loss due to the disease and ageing. This model was applied to COVID-19 outbreak data from different countries as well as the role of comorbidities and sensitivity of the model to various parameters was established.

- We applied the ideas in the Boolean dynamic context to comorbidity risk factor (obesity) by establishing a modeling epidemic propagation in a Boolean framework (Hop field-type propagation equation, notion of centrality to interpret interaction graphs, existence of stationary states), we presented an application of the model chosen within the framework of the observation of real data and we show the contribution of the model in their interpretation in order to increase the efficiency of the model currently used, for the understanding of the mechanisms and the prediction of the dynamic behaviours of the pandemics present and future.

5.2 General Conclusion and Some Perspectives

We have provided in this thesis the intertwine between mathematical and statistical models for the modeling of infectious disease.
Concerning the contagious diseases, public health physicians and policy-makers are constantly faced with four challenges. The first concerns the estimation of the basic reproduction number R_0. The systematic use of R_0 simplifies the decision-making process by policy-makers, advised by public health authorities, but it is too caricature to account for the biology behind the viral spread. We have observed that R_0 was not constant during an epidemic wave due to exogenous and endogenous factors influencing both the duration of the contagiousness period and the transmission rate during this phase. Then, the first challenge concerns the estimation of the mean duration of the contagiousness period for infected patients. As for the transmission rate, realistic assumptions made it possible to obtain an upper limit to this duration, in order to better guide the individual quarantine or lockdown measures decided by the authorities in charge of public health. This upper bound also makes it possible to obtain a lower bound for the percentage of unreported infected patients, which gives an idea of the quality of the census of cases of infected patients, which is the second challenge facing specialists of contagious diseases. The third challenge is the estimation of the daily reproduction numbers over the contagiousness period and the fourth and final interesting

challenge is the extension of the methods developed in the present paper to contagious non-infectious diseases (i.e., those without causal infectious agents), such as social contagious diseases, the best example being that of the pandemic linked to obesity, for which many concepts and modelling methods presented here remain available.
Also, the severity of public health mitigation strategies implemented in different cities or regions across countries demonstrates that demographic fluctuations are important factors for the spread and prevalence of COVID-19 due to its stochastic transmission nature. Environmental factors have an impact on these demographic events. We discovered that demographic factors, regardless of reopening or closing strategies, can change the time and scale of infection. The public health mitigation, in conjunction with our mathematical and statistical models of the COVID-19 pandemic, is used to build a detailed picture of the distribution of immunity to SARS-CoV-2 and determine the severity of consecutive waves of infection given the specific public health mitigation programs. Test, trace and isolate approaches will better control the disease spread but observations from different countries showed how limited capacities for this approach can imply a fast growth in undetected incidence cases. Adaptions of non-pharmaceutical interventions were associated with changes and fluctuation in the effective reproduction number and also in the daily reproduction rates our modeling approach has calculated in this thesis. The perspective generated by studying epidemic ageing model in human populations support the continued use of matrix population models for understanding the epidemiology of viral infections across the life span, by noticing that each time the virus mutates, it is necessary to re-estimate the coefficients of the matrix model again.
Finally, we have presented a large scale data analysis and also assessed the efficiency of many statistical models. We have adopted a retrospective modelling approach which has shown to us that the results are coherent with observations based on different parameters or counts in the COVID-19 pandemic having in mind that the prevention protocols varied during the different phases of the pandemic throughout the epidemic period. Interestingly, the results of the prediction models show these models are more or less adapted for COVID-19 data set with different fluctuations. Through the fitting and prediction of the data set, we reveal that some of the models have more advantages in the prediction of the epidemiological data coming from COVID-19 outbreak.

5.3 Recommendations

While previous research by our research network indicated that geo-climatic factors such as temperature change is likely to extend the geographical spread of the disease and has had the greatest influence on the spread of the epidemic, we were able to demonstrate that socio-economic factors such as GDP dedicated to health expenditures and demographic factors such as age also had the greatest influence on the spread of the epidemic. We observe a correlation of about 42 percent when GDP dedicated to health expenditures is correlated with epidemiologic variable, and of course, using ordinary least square method by adding more socioeconomic variable to the modeling, a strong correlation is observed, proving that GDP dedicated to health expenditures takes a chunk of this correlation. In this thesis, we used different modeling strategies to demonstrate the impact of vaccination on older and younger age groups, as well as to demonstrate a faster growth of infectious in the younger sub-population than the older sub-population since younger population with active social activities are more likely to be exposed to COVID-19. Our modeling approach was able to

confirm what was observed in real life, indicating that age has an impact on epidemic spread. The role of vaccination has been globally positive, most notably in the reduction of mortality and hospitalization cases, which we studied using a functional data analysis approach to study the dynamics of diseases in French departments before and after vaccination. As a result, it is critical to encourage people to get vaccinated in order to avoid severe illness and hospitalization. One key reason we predicted the test response variable using functional linear model is because multiple testing campaigns helps to improve COVID-19 surveillance day by day, particularly in the case of "false" peaks where one observe only shoulders detected by our inflection point estimation method, reason we recommend that researchers and modelers should not only estimate parameters or be interested in the dynamics of the disease at the exponential phase alone but also at the shoulders before the different waves of the pandemic which we have investigated in this thesis since many countries don't attain the different peaks detected during the spread of the disease.
We also observed that socio-economic demographic variables show an anti-correlation for vaccination (Dose 1 and Dose 2) and number of tests. This demonstrates that individuals with a high socio-cultural level are more likely to monitor their health by getting tested and vaccinated when vaccines were introduced because these socioeconomic factors we considered affects the quality of life.
We investigated all of these in order to localize mitigation measures and how to allocate intervention resources for future epidemics using sub-population modeling approach, and if government policies can aid in the increase in GDP dedicated to health expenditures, allowing countries to strengthen their health infrastructures, it will aid in proper preparedness for future pandemics and preventing future pandemic outbreaks because the COVID-19 epidemic has wreaked havoc on the healthcare system as well as the socioeconomic standing of many countries.
These modeling approaches we have presented will be useful for predicting, preventing, and controlling the spread of COVID-19, and thus future epidemics.

5.4 Future Work

There are several possible extensions of the two approaches in this thesis.
On the mathematical modeling side, in the future, we will compare the results generated from the estimation of daily reproduction rates by varying contagiousness duration between 3 and 12 days in order to obtain possibly more realistic values for the R_j's, and hence, have perhaps a double explanation for the 7 days periodicity, both sociological and biological. One interesting challenge for infectious disease modeling is the improvement and extension of the decovolution method developed here to the contagious non-infectious diseases (i.e., without causal infectious agent), such as social contagious diseases, the best example being that of the pandemic linked to obesity, for which many concepts and modelling methods remain available and our future research will propose new avenues of research in this field. In the first case (discrete), the two-state variable "vaccinated or not" should be added, in order to properly take into account the effect of the vaccination policy. We intend to extend the epidemic ageing model in human population to more age classes (6, 7, etc.) in more countries in which data are available in order to better understand the evolution of the disease in subpopulations; now that we are observing a sharp increase in the number of children infected and those who interact with them, such as parents and teachers in some countries; then, we will consider the role of comorbidities in the subpopulations of more

countries, and build a global continuous model. It is possible to use a matrix population model that considers Intensive Care Unit (ICU) and hospitalization cases due to COVID-19 and then see the effect of therapies at different life history points. Another future work for age dependent modeling should, for example, take into account more age groups, at least four: children (age $\leq$ 12 years), adolescents (12 years $<$ age $\leq$ 18 years), young adults (18 years $<$ age $\leq$ 65 years), and older adults ($>$65 years). The most important pitfall in simulations of such a model is the estimation (by observation, calculation, or assumption) of its parameters, already difficult with two age groups. A random choice of the values of the parameters in plausible intervals followed by a study of the sensitivity to the parameters of the model, could make it possible to partly overcome the constraint of parameter estimation in a more precise future epidemiological–demographic model, in particular with regard to age groups. Also we can introduce an accelerated ageing due to infection, as well as an influence of exogenous determinants such as geo-climatic, socio-economic, and health-related factors, which weight differently on the different age groups, therefore, change the growth dynamics specific to each of the sub-populations studied in this thesis. It is possible to estimate cell ageing due to the virulence which supplementary cell lifespan loss which could cause pathological events like infectious diseases, by using discrete and continuous formulation of Hahn's model. From endotracheal cytometry we count the alveolar cells in order to have information about their loss and also from the protein profile we will get information on proteins involved in apoptosis like Gaf2 and then this information will be fielded into the mathematical model (Hahn's model) in order to estimate the lifespan loss. Also, we can introduce the influence of altitude in the modeling of the cell lifespan loss because it has been established that the beneficial effects of altitude on lung function could be due to changes in the expression level of ACE2, which is not only the primary route of entry for the COVID-19 virus into cells, but also appears to play a key role in acute lung injury and other organ damages.
On the statistical modeling side, to benchmark the statistical studies in this thesis, mixed effect models, generalized additive models and spatial data analysis will be used to further predict the evolution and dynamics of the spread of COVID-19 pandemic. As demonstrated by the COVID-19 pandemic, spatial epidemiology is an important field of study. For instance, Hidden Markov models are popular time series models. It can be defined over discrete or continuous time. In the mathematical modeling in this thesis, we have been able to develop an epidemic ageing model for human populations which we can now exploit using a statistical approach by studying the State-dependent distributions and state estimations, developing transition probability matrix, marginal distributions and temporal dependence, estimating the initial state distribution and assessing model adequacy using forecast (pseudo-)residuals and posterior predictive checks. Also, spatial statistics and disease mapping are well-developed in literature, however, there are notable computational bottlenecks when it comes to the application of spatially-fine grained epidemiological models. We will investigate the application of recently proposed deep learning approaches to efficiently encode spatiotemporal data before standard (non-spatial) epidemiological modeling approaches, such as SEIR, renewal process, and ODE-based models. Spatial epidemiology is designed to find patterns, and in this case, it will help us detect peaks at different waves of the pandemic while taking demographic factors into account. The goal is to investigate the evolution and dynamics of disease spread over time along the epidemic curve in various regions, countries, and continents. Also, we can investigate the relationship between comorbidities like obesity, diabetes, cardiovascular diseases or asthma and epidemic-socio-demo-economic determinants, using several machine learning and deep learning methods and then predict some of these determi-

nants since we have already shown in this thesis how they were associated with COVID-19 prevalence. In addition to improving the section of the thesis dedicated to spectral analysis methods in which time (days) was converted to frequency, it is possible to examine the periodicity and frequency peaks of time series data by converting time series data (pandemic) from a time domain to a frequency domain and applying a Dense Convolutional Neural Network (DenseCNN) or CNN and producing the scalogram generated from Discrete Fourier transform (DFT) transformation of the pandemic data and training the DenseCNN for a time series by performing a regression or classification target.
Also, another future work could be to try to model the change-points between different phases of a pandemic, namely the transition between an endemic phase and the future epidemic wave. The estimation and forecast of this transition time is a difficult challenge rarely faced by the modelers (using in general change-points estimation in linear models or algorithms for detection of abrupt changes in signals and systems) and could constitute if it provides accurate and reliable methods a huge progress in epidemic studies.
Finally, the population distribution of viral loads observed under random or symptom-based surveillance reverse transcription quantitative polymerase chain reaction (RT-qPCR) tests, which provide semiquantitative results in the form of cycle threshold (Ct) values that change during an epidemic and are an important predictor of pandemic trajectory, should be investigated. We intend to estimate the epidemic's trajectory using both statistical modeling (in this case, the branching process) and mathematical modeling (in this case, the exponential model and the SEIR model).
The publications in preparation are as follows:

- K. Oshinubi, J.P. Olumuyiwa, M. Rachdi and J. Demongeot. The modeling of COVID-19 data sets using generalized addictive models and mixed models (in preparation).
- K. Oshinubi, M. Rachdi and J. Demongeot. Spatial Epidemiology Modeling of COVID-19 Pandemic (in preparation).
- K. Oshinubi, S. Flaxman, M. Rachdi and J. Demongeot. Estimating Viral Load Distribution using a Time-varying Branching Process and Epidemic Transmission Models (in preparation).

Appendix A

Appendix

Table A.1: Presentation of the Pearson correlation coefficients between 20 numbers of world daily new cases observed between the days 34 to 53 after the 24 January 2020 (date of the start of the COVID-19 outbreak with confirmed cases in Europe) and series of 20 numbers of world daily new cases observed in running windows of length 20 days until day 213, calculated using the online facility https://www.socscistatistics.com/tests/pearson/default2.aspx, and data from [119]. The last column of the Table gives the mean of the correlation coefficients between the first window and the 20 running windows starting at week day E and after the number of such 20 correlation coefficients which are significantly strictly positive.

Day N	**Daily cases**	**PCN**	**D**	**PCD**	**E**	**W**	**PCEW**	**MCC**
34, Wed	1155							
35, Thu	1472							
36, Fri	1835							
37, Sat	2218							
38, Sun	2662							
39, Mon	2644							
40, Tue	3331							
41	3206							
42	4327							
43	5174							
44	5771							
45	5963							
46, Mon	7294							
47	8858							
48	11465							
49	13420							
50	16767							
51	17329							
52	19500							
53, Mon	20715							
54, Tue	22788	0.871903	1	0.871903	1	1	0.871903	0.513784 14/20
55, Wed	26840	0.833092	2	0.833092	1	2	0.665054	
56, Thu	31636	0.798643	3	0.798643	1	3	0.052967	
57, Fri	38016	0.774313	4	0.774313	1	4	0.215606	
58, Sat	33019	0.763159	5	0.763159	1	5	0.296313	

Day N	Daily cases	PCN	D	PCD	E	W	PCEW	MCC
59, Sun	36073	0.789508	6	0.789508	1	6	0.395235	
60, Mon	43671	0.767306	7	0.767306	1	7	0.428514	
61	44250	0.665054	1	0.665054	1	8	0.669878	
62	48363	0.540774	2	0.540774	1	9	0.778017	
63	59111	0.383754	3	0.383754	1	10	0.719714	
64	62936	0.311278	4	0.311278	1	11	0.625795	
65	64624	0.266571	5	0.266571	1	12	0.66546	
66	57929	0.276242	6	0.276242	1	13	0.780229	
67	61794	0.230716	7	0.230716	1	14	0.72027	
68	70581	0.052967	1	0.052967	1	15	0.707716	
69	74667	-0.06834	2	-0.06834	1	16	0.653118	
70	76882	-0.17189	3	-0.17189	1	17	0.656816	
71	81598	-0.17193	4	-0.17193	1	18	0.368153	
72	78609	-0.00811	5	-0.00811	1	19	0.004649	
73	68517	0.311795	6	0,311795	1	20	0.000278	
74	71938	0.39616	7	0.39616	2	1	0.833092	0.383302 13/20
75	76716	0.215606	1	0.215606	2	2	0.540774	
76	82115	0.027762	2	0.027762	2	3	-0.06834	
77	83995	-0.03645	3	-0.03645	2	4	0.027762	
78	90241	-0.05802	4	-0.05802	2	5	0.248403	
79	77692	0.116351	5	0.116351	2	6	0.145776	
80	70237	0.307003	6	0.307003	2	7	0.389544	
81	69497	0.298721	7	0.298721	2	8	0.545406	
82	71977	0.296313	1	0.296313	2	9	0.71472	
83	78993	0.248403	2	0.248403	2	10	0.571106	
84	79123	0.202059	3	0.202059	2	11	0.56017	
85	81357	0.320394	4	0.320394	2	12	0.609405	
86	78838	0.440548	5	0.440548	2	13	0.718404	
87	72258	0.527497	6	0.527497	2	14	0.602925	
88	73597	0.526067	7	0.526067	2	15	0.605019	
89	73426	0.395235	1	0.395235	2	16	0.494937	
90	77928	0.145776	2	0.145776	2	17	0.477868	
91	82020	0.08347	3	0.08347	2	18	0.071667	
92	96755	0.102804	4	0.102804	2	19	-0.18503	
93	87252	0.320819	5	0.320819	2	20	-0.23757	
94	71139	0.499389	6	0.499389	3	1	0.798643	0.346759 13/20
95	67460	0.548743	7	0.548743	3	2	0.383754	
96	74207	0.428514	1	0.428514	3	3	-0.17189	
97	76455	0.389544	2	0.389544	3	4	-0.03645	
98	84890	0.400223	3	0.400223	3	5	0.202059	
99	92685	0.470877	4	0.470877	3	6	0.08347	
100	81538	0.632728	5	0.632728	3	7	0.400223	
101	81295	0.699566	6	0.699566	3	8	0.451569	
102	78898	0.715578	7	0.715578	3	9	0.698072	
103	79759	0.669878	1	0.669878	3	10	0.527833	
104	93179	0.545406	2	0.545406	3	11	0.574066	
105	93583	0.451569	3	0.451569	3	12	0.621132	

Day N	Daily cases	PCN	D	PCD	E	W	PCEW	MCC
106	93736	0.506118	4	0.506118	3	13	0.674633	
107	86990	0.607988	5	0.607988	3	14	0.59476	
108	78677	0.702738	6	0.702738	3	15	0.575404	
109	71674	0.771323	7	0.771323	3	16	0.488039	
110	84952	0.778017	1	0.778017	3	17	0.402127	
111	88726	0.71472	2	0.71472	3	18	0.009761	
112	96296	0.698072	3	0.698072	3	19	-0.11507	
113	99237	0.692351	4	0.692351	3	20	-0.22695	
114	95176	0.726056	5	0.726056	4	1	0.774313	0.38065 12/20
115	81646	0.75754	6	0.75754	4	2	0.311278	
116	90580	0.783995	7	0.783995	4	3	-0.17193	
117	95791	0.719714	1	0.719714	4	4	-0.05802	
118	103269	0.571106	2	0.571106	4	5	0.320394	
119	108305	0.527833	3	0.527833	4	6	0.102804	
120	107714	0.562433	4	0.562433	4	7	0.470877	
121	100380	0.610005	5	0.610005	4	8	0.506118	
122	97159	0.666455	6	0.666455	4	9	0.692351	
123	91816	0.686601	7	0.686601	4	10	0.562433	
124	93546	0.625795	1	0.625795	4	11	0.602733	
125	107449	0.56017	2	0.56017	4	12	0.667945	
126	117149	0.574066	3	0.574066	4	13	0.685452	
127	126792	0.602733	4	0.602733	4	14	0.594878	
128	125224	0.634611	5	0.634611	4	15	0.592655	
129	109948	0.746575	6	0.746575	4	16	0.59711	
130	105959	0.788147	7	0.788147	4	17	0.462685	
131	116523	0.66546	1	0.66546	4	18	0.043801	
132	121422	0.609405	2	0.609405	4	19	0.000826	
133	131498	0.621132	3	0.621132	4	20	-0.1457	
134	131555	0.667945	4	0.667945	5	1	0.763159	0.467941 13/20
135	129557	0.726055	5	0.726055	5	2	0.266571	
136	114896	0.795896	6	0.795896	5	3	-0.00811	
137	108855	0.811018	7	0.811018	5	4	0.116351	
138	122083	0.780229	1	0.780229	5	5	0.440548	
139	136934	0.718404	2	0.718404	5	6	0.320819	
140	139155	0.674633	3	0.674633	5	7	0.632728	
141	143260	0.685452	4	0.685452	5	8	0.607988	
142	135048	0.733599	5	0.733599	5	9	0.726056	
143	124422	0.777163	6	0.777163	5	10	0.610005	
144	126372	0.784775	7	0.784775	5	11	0.634611	
145	144280	0.72027	1	0.72027	5	12	0.726055	
146	146482	0.602925	2	0.602925	5	13	0.733599	
147	141526	0.59476	3	0.59476	5	14	0.668216	
148	182683	0.594878	4	0.594878	5	15	0.667777	
149	158002	0.668216	5	0.668216	5	16	0.683387	
150	131120	0.736841	6	0.736841	5	17	0.537366	
151	140363	0.759236	7	0.759236	5	18	0.114956	
152	164750	0.707716	1	0.707716	5	19	0.161079	

Day N	Daily cases	PCN	D	PCD	E	W	PCEW	MCC
153	174539	0.605019	2	0.605019	5	20	-0.04435	
154	181353	0.575404	3	0.575404	6	1	0.789508	0.557804 14/20
155	195611	0.592655	4	0.592655	6	2	0.276242	
156	178369	0.667777	5	0.667777	6	3	0.311795	
157	164816	0.706037	6	0.706037	6	4	0.307003	
158	162497	0.692452	7	0.692452	6	5	0.527497	
159	175710	0.653118	1	0.653118	6	6	0.499389	
160	199422	0.494937	2	0.494937	6	7	0.699566	
161	210308	0.488039	3	0.488039	6	8	0.702738	
162	215135	0.59711	4	0.59711	6	9	0.75754	
163	196362	0.683387	5	0.683387	6	10	0.666455	
164	178891	0.7602	6	0.7602	6	11	0.746575	
165	172555	0.783989	7	0.783989	6	12	0.795896	
166	209690	0.656816	1	0.656816	6	13	0.777163	
167	215081	0.477868	2	0.477868	6	14	0.736841	
168	225012	0.402127	3	0.402127	6	15	0.706037	
169	238630	0.462685	4	0.462685	6	16	0.7602	
170	217665	0.537366	5	0.537366	6	17	0.600851	
171	199356	0.600851	6	0.600851	6	18	0.195316	
172	197440	0.574631	7	0.574631	6	19	0.282929	
173	221056	0.368153	1	0.368153	6	20	0.016546	
174	236580	0.071667	2	0.071667	7	1	0.767306	0.571253 15/20
175	250058	0.009761	3	0.009761	7	2	0.230716	
176	241680	0.043801	4	0.043801	7	3	0.39616	
177	227333	0.114956	5	0.114956	7	4	0.298721	
178	222884	0.195316	6	0.195316	7	5	0.526067	
179	206535	0.208246	7	0.208246	7	6	0.548743	
180	240767	0.004649	1	0.004649	7	7	0.715578	
181	281499	-0.18503	2	-0.18503	7	8	0.771323	
182	276640	-0.11507	3	-0.11507	7	9	0.783995	
183	289991	0.000826	4	0.000826	7	10	0.686601	
184	261967	0.161079	5	0.161079	7	11	0.788147	
185	225286	0.282929	6	0.282929	7	12	0.811018	
186	219632	0.258542	7	0.258542	7	13	0.784775	
187	249892	0.000278	1	0.000278	7	14	0.759236	
188	290363	-0.23757	2	-0.23757	7	15	0.692452	
189	287597	-0.22695	3	-0.22695	7	16	0.783989	
190	291504	-0.1457	4	-0.1457	7	17	0.574631	
191	259867	-0.04435	5	-0.04435	7	18	0.208246	
192	224589	0.016546	6	0.016546	7	19	0.258542	
193	200970	0.038824	7	0.038824	7	20	0.038824	
194	255117							
195	272913							
196	283977							
197	283782							
198	271031							
199	225645							

Day N	Daily cases	PCN	D	PCD	E	W	PCEW	MCC
200	219202							
201	268712							
202	287955							
203	286629							
204	287391							
205	265682							
206	216826							
207	198498							
208	254103							
209	272134							
210	267720							
211	258578							
212	268958							
213	210600							

Note: PCN is Pearson correlation coefficient of the 140 mobile windows starting at day N, PCD is Pearson correlation coefficient of the 20 mobile windows starting at week day D, PCEW is Pearson correlation coefficient of the 140 mobile windows starting at week day E of the week W, and MCC is Mean of the correlation coefficients between the first window and the 20 mobile windows starting at week day E and number of such correlation coefficients significantly > 0.

Table A.2: Calculation of the daily R_j's and shape of their distribution for 194 countries and for the two first waves.

Country Name	First wave			Second wave		
	R_0	R_j's	U-shape	R_0	R_j's	U-shape
AFGHANISTAN	0.65	0.17; 0.09; 0.39	YES	0.04	−1.38; −0.36; 1.78	INCR
ALGERIA	1.25	3.93; −6.21; 3.53	YES	0.91	1.28; −1.06; 0.69	YES
ARUBA	5.46	10.31; −39.32; 34.47	YES	1.10	1.54; −1.60; 1.16	YES
ANDORRA	1.36	1.00; 0.79; −0.43	DECR	0.12	4.34; −1.63; −2.59	DECR
ANGOLA	0.63	0.33; 1.42; −1.12	INV	1.70	9.22; −1.58; −5.94	DECR
ANTIGUA	1.92	0.00; 1.25; 0.67	INV	2.13	−0.40; 1.33; 1.20	INV
ALBANIA	0.96	0.48; 0.50; −0.02	INV	0.66	1.98; −0.56; −0.76	DECR
ARGENTINA	0.73	0.57; −1.28; 1.44	YES	0.36	1.27; 0.75; −1.66	DECR
ARMENIA	4.43	17.99; −36.99; 23.43	YES	0.86	1.41; −0.97; 0.42	YES
AUSTRALIA	2.79	−1.02; 3.47; 0.34	YES	1.50	−0.88; 0.68; 1.70	INCR
AUSTRIA	1.17	−1.78; −0.05; 3.00	INCR	2.08	0.62; −3.55; 5.01	YES
AZERBAIJAN	1.16	1.23; −1.32; 1.25	YES	0.37	10.36; −6.45; −3.54	YES
BAHAMAS	0.57	−0.13; −0.98; 1.68	YES	1.22	0.22; −0.86; 1.86	YES
BAHRAIN	1.10	−0.74; 0.28; 1.56	INCR	1.14	1.98; −2.69; 1.85	YES
BANGLADESH	1.04	2.37; −2.97; 1.64	YES	0.99	0.86; −0.69; 0.82	YES
BARBADOS	1.86	0.86; −0.64; 1.64	YES	1.14	0.22; −0.81; 1.73	YES
BELARUS	1.57	−2.37; −4.58; 8.52	YES	1.07	−0.33; 0.24; 1.16	INCR
BELGIUM	0.43	11.66; −15.63; 4.41	YES	2.23	1.17; −2.39; 3.45	YES
BELIZE	0.99	0.80; 0.42; −0.23	DECR	0.51	1.77; −0.21; −1.05	DECR
BENIN	0.85	0.81; 0.47; −0.43	DECR	0.85	1.17; 0.22; −0.54	DECR

Country Name	First wave			Second wave		
	R_0	R_j's	U-shape	R_0	R_j's	U-shape
BHUTAN	15.00	14.00; 15.00; −14.00	INV	1.08	0.80; 0.57; −0.29	DECR
BOLIVIA	2.17	8.47; −1.17; −5.13	DECR	1.61	0.96; −0.30; 0.95	YES
BOSNIA	0.09	−1.06; −1.05; 2.20	INCR	1.56	−0.57; −0.51; 2.64	INCR
BOTSWANA	28.47	0.22; 0.00; 28.25	YES	28.43	0.22; −0.05; 28.26	YES
BRAZIL	0.77	0.31; 1.08; −0.62	INV	0.46	1.21; 0.16; −0.91	DECR
BRUNEI	1.08	0.10; −0.15; 1.13	YES	1.00	1.00; −1.00; 1.00	YES
BULGARIA	5.06	14.73; −66.02; 56.35	YES	0.75	1.34; −0.98; 0.39	YES
BURKINA FASO	1.08	0.72; −0.34; 0.70	YES	0.94	0.31; 0.24; 0.39	YES
BURUNDI	1.33	1.33; −0.67; 0.67	YES	2.18	0.53; 1.80; −0.15	INV
CABO VERDE	0.82	−0.08; −0.26; 1.16	YES	0.19	0.56; 1.37; −1.74	INV
CAMBODIA	0.34	0.08; 0.25; 0.01	INV	0.27	0.06; 0.15; 0.06	INV
CAMEROON	2.17	2.36; 1.25; −1.44	DECR	2.48	0.50; −0.25; 2.23	YES
CANADA	1.10	−0.55; −0.72; 2.37	YES	0.44	2.36; −0.44; −1.48	DECR
CAR	1.66	−0.07; 0.64; 1.09	INCR	0.33	0.44; −0.22; 0.11	YES
CHAD	1.19	0.77; −1.15; 1.57	YES	0.77	1.19; 0.25; −0.67	DECR
CHILE	1.00	0.72; 0.17; 0.11	DECR	1.64	0.37; −4.45; 5.72	YES
CHINA	1.10	0.90; −0.49; 0.69	YES	0.87	1.16; 0.60; −0.89	DECR
COLUMBIA	1.00	1.75; −0.86; 0.11	YES	1.47	−1.14; 3.08; −0.47	INV
COMOROS	3.75	0.00; −2.75; 6.5	YES	1.65	−0.58; 1.24; 0.99	INV
CONGO DEM.	0.03	−0.37; −0.39; 0.79	YES	0.88	0.66; 0.74; −0.52	INV
CONGO REP.	0.92	0.92; 0.92; −0.92	DECR	0.39	−0.12; 0.19; 0.32	INCR
COSTA RICA	0.50	−2.79; −3.84; 7.13	YES	1.26	1.21; −0.85; 0.90	YES
COTE D'VOIRE	1.18	−0.49; −0.63; 2.30	YES	2.09	4.32; −7.09; 4.86	YES
CROTIA	0.75	0.53; 0.79; −0.57	INV	0.57	0.68; −0.64; 0.53	YES
CUBA	0.48	−37.25; 16.17; 21.56	INCR	0.78	0.34; −0.73; 1.17	YES
CURACAO	0.50	3.00; −1.00; −1.50	DECR	4.19	1.93; −4.01; 6.27	YES
CYPRUS	0.69	0.27; 2.49; −2.07	INV	0.45	−0.42; 1.76; −0.89	INV
CZECH	0.16	−0.16; 3.88; −3.56	INV	0.88	1.88; −1.41; 0.41	YES
DENMARK	0.80	−0.11; 0.41; 0.50	INCR	0.64	−0.03; 4.65; −3.98	INV
DJIBOUTI	0.17	1.23; 0.24; −1.30	DECR	0.36	0.64; 0.41; −0.69	DECR
DOMINICAN	1.02	1.05; −0.31; 0.28	YES	1.57	0.32; −0.06; 1.31	YES
DOMINICA	7.75	2.00; −4.00; 9.75	YES	0.67	−0.36; 0.72; 0.31	INV
ECUADOR	1.46	−0.47; 1.06; 0.87	INV	1.14	0.73; −0.14; 0.55	YES
EGYPT	0.84	0.30; 0.37; 0.17	INV	0.51	11.99; −3.76; −7.72	DECR
EL SALVADOR	1.70	−0.20; 0.59; 1.31	INCR	0.66	−0.76; −14.49; 15.91	YES
EQUATORIAL G.	0.38	0.85; −0.20; −0.27	DECR	1.48	0.81; −0.66; 1.33	YES
ERITREA	1.18	1.44; −0.05; −0.21	DECR	0.80	1.02; 0.20; −0.42	DECR
ESTONIA	0.87	1.96; 0.82; −1.91	DECR	3.04	−0.70; −1.80; 5.54	YES
ESWATINI	0.94	1.41; −1.42; 0.95	YES	0.71	−0.02; 1.52; −0.79	INV
ETHIOPIA	0.80	−0.56; −1.45; 2.81	YES	1.24	0.34; 0.13; 0.77	YES
FIJI	2.00	0.00; 1.00; 1.00	INCR	0.50	0.75; −0.50; 0.25	YES
FINLAND	1.14	0.91; −0.42; 0.65	YES	2.41	0.56; −2.38; 4.23	YES
FRANCE	1.17	0.82; 0.10; 0.25	YES	2.17	0.88; −0.86; 2.15	YES
GABON	0.97	0.20; 0.47; 0.30	INV	0.19	−0.51; 0.00; 0.70	INCR
GAMBIA	0.83	−0.25; 0.43; 0.65	INCR	0.37	−0.38; 0.00; 0.75	INCR
GEORGIA	1.23	0.16; 0.43; 0.64	INCR	0.79	1.52; −0.49; −0.24	YES

Country Name	First wave			Second wave		
	R_0	R_j's	U-shape	R_0	R_j's	U-shape
GERMANY	0.73	0.15; −1.04; 1.62	YES	0.79	1.15; −0.56; 0.20	YES
GHANA	1.48	0.55; 0.70; 0.23	INV	0.62	0.13; −0.81; 1.30	YES
GREECE	0.71	0.33; −0.27; 0.65	YES	0.71	0.95; 0.28; −0.52	DECR
GRENADA	14.00	−5.00; 3.00; 16.00	INCR	0.10	−0.15; 0.00; 0.25	INCR
GUADELOUPE	1.35	0.00; 0.76; 0.59	INV	1.35	0.00; 0.76; 0.59	YES
GUATEMALA	0.25	2.01; −0.70; −1.06	YES	0.27	1.19; −0.11; −0.81	DECR
GUIANA	0.88	1.30; −0.38; −0.04	YES	0.43	0.99; 0.27; −0.83	DECR
GUINEA	0.46	0.65; −0.56; 0.37	YES	1.68	0.21; 0.68; 0.79	INCR
GUINEA BISSAU	1.14	0.06; 1.59; −0.51	INV	4.20	−0.11; 0.04; 4.27	INCR
GUYANA	2.38	−3.45; −0.20; 6.03	INCR	4.23	−0.53; 0.58; 4.18	INCR
HAITI	0.60	0.30; −0.13; 0.43	YES	0.61	0.32; 0.42; −0.13	INV
HONDURAS	0.57	−2.94; 3.12; 0.39	INV	1.64	0.13; 0.54; 0.97	INCR
HONG KONG	0.04	0.95; −0.69; −0.22	YES	0.24	2.50; −8.79; 6.53	YES
HUNGARY	0.90	0.66; −0.12; 0.36	YES	1.93	1.91; −2.72; 2.74	YES
ICELAND	2.28	−0.85; 3.93; −0.80	INV	0.66	0.84; 0.22; −0.40	NO
INDIA	0.98	1.82; 0.53; −1.37	DECR	0.96	1.08; −0.57; 0.45	YES
INDONESIA	0.95	0.67; 0.88; −0.60	INV	0.99	1.06; −0.03; −0.03	YES
IRAN	1.04	1.73; −0.67; −0.02	YES	0.90	6.62; −6.62; 0.90	YES
IRAQ	0.77	0.15; −0.35; 0.96	YES	0.96	0.77; −0.40; 0.59	YES
IRELAND	2.16	−2.83; −5.64; 10.63	YES	1.12	1.12; −0.39; 0.39	YES
ISRAEL	0.21	−1.39; 1.08; 0.52	INV	1.16	−0.16; 0.44; 0.88	INCR
ITALY	1.04	2.24; −1.85; 0.65	YES	3.69	1.65; −7.89; 9.93	YES
JAMAICA	0.43	0.13; 0.06; 0.24	YES	2.47	−0.34; 2.06; 0.75	INV
JAPAN	1.02	0.69; 0.88; −0.55	INV	1.16	0.61; 0.42; 0.13	DECR
JORDAN	2.53	10.82; −18.20; 9.91	YES	0.93	1.28; 0.57; −0.92	DECR
KAZAKHSTAN	0.60	0.53; −5.45; 5.52	YES	2.06	−0.05; 2.37; −1.26	INV
KENYA	1.14	0.05; 0.65; 0.44	INV	1.18	0.47; 1.34; −0.63	INV
KOREA REP.	1.00	0.12; 0.87; 0.01	INV	1.04	0.60; −0.03; 0.47	YES
KOSOVO	1.02	1.00; 1.02; −1.00	INV	0.99	1.31; −0.29; −0.03	YES
KUWAIT	0.88	0.5; −0.34; 0.67	YES	1.10	0.58; −0.84; 1.36	YES
KYRGYZSTAN	0.17	−0.73; 0.26; 1.64	INCR	1.05	0.28; −0.32; 1.09	YES
LAO PDR	0.50	0.50; 0.50; −0.50	DECR	0.15	0.33; 0.74; −0.92	INV
LATVIA	0.74	1.97; −0.76; −0.47	YES	0.50	0.40; −0.22; 0.32	YES
LEBANON	1.03	0.57; 0.12; 0.34	YES	0.90	0.23; 0.06; 0.61	YES
LESOTHO	7.08	−2.86; 7.22; 2.72	INV	1.42	0.37; 1.51; −0.46	INV
LIBERIA	0.31	0.18; −0.04; 0.17	YES	4.56	0.14; 4.61; −0.19	INV
LIBYA	0.96	0.19; −0.71; 1.48	YES	0.79	−0.42; 0.56; 0.65	INCR
LITHUANIA	0.83	0.56; 0.11; 0.16	YES	2.49	−0.90; −0.52; 3.91	INCR
LUXEMBOURG	0.24	−8.55; −3.75; 12.54	INCR	1.48	1.16; −0.91; 1.23	YES
MACAO	0.29	1.14; 2.29; −3.14	INV	-	-	-
MADAGASCAR	0.94	0.61; −0.16; 0.49	YES	0.75	0.38; −1.54; 1.91	YES
MALAWI	1.12	−0.23; 0.53; 0.82	INCR	6.46	−0.41; 0.99; 5.88	INCR
MALAYSIA	1.25	0.38; 2.79; −1.92	INV	1.30	−0.57; 1.82; 0.05	INV
MALDIVES	0.83	0.60; −0.53; 0.76	YES	1.05	−0.27; 0.70; 0.62	INV
MALI	0.64	0.59; 0.42; −0.37	DECR	7.78	−2.64; −4.96; 15.38	YES
MALTA	1.06	1.15; 0.24; −0.33	DECR	0.99	−0.73; 1.81; −0.09	INV

Country Name		**First wave**			**Second wave**	
	R_0	R_j**'s**	**U-shape**	R_0	R_j**'s**	**U-shape**
MAURITANIA	1.76	−0.94; 0.29; 2.41	INCR	1.14	0.73; −0.41; 0.82	YES
MAURITIUS	4.49	−4.05; 0.36; 8.18	INCR	0.35	1.41; 0.53; −1.59	DECR
MAYOTTE	5.46	−9.46; −2.50; 17.42	INCR	1.05	0.72; −0.17; 0.50	YES
MEXICO	0.86	−1.39; 3.07; −0.82	INV	2.53	−0.55; 0.10; 2.98	INCR
MOLDOVA	1.03	2.73; −0.67; −1.03	DECR	0.36	1.27; 0.66; −1.57	DECR
MONACO	3.15	0.52; −1.93; 4.56	YES	0.54	1.02; −0.12; −0.36	DECR
MONGOLIA	10.25	1.25; 19.25; −10.25	INV	0.68	0.91; 0.25; −0.48	DECR
MONTENEGRO	1.37	2.94; −3.90; 2.33	YES	0.66	2.36; 0.26; −1.96	DECR
MOROCCO	0.90	0.36; 1.41; −0.87	INV	0.95	0.95; −0.15; 0.15	YES
MOZAMBIQUE	0.72	0.92; 0.001; −0.20	DECR	0.70	2.46; −2.45; 0.69	YES
MYANMAR	1.12	−0.75; 1.07; 0.80	INV	1.15	−1.36; −2.17; 4.68	YES
NAMIBIA	0.68	1.37; −1.82; 1.13	YES	1.22	−0.26; 0.95; 0.53	INV
NEPAL	0.74	0.35; 0.76; −0.37	INV	0.78	0.11; 0.58; 0.09	INV
NETHERLANDS	1.19	0.11; 0.11; 0.97	YES	1.04	1.05; −0.99; 0.98	YES
CALEDONIA	5.00	−2.00; 2.00; 5.00	YES	1.00	1.00; −1.00; 1.00	YES
NEW ZEALAND	0.74	2.30; −3.40; 1.84	YES	0.72	−0.52; 0.43; 0.81	INCR
NICARAGUA	0.97	−0.03; 0.97; 0.03	INV	1.02	0.86; 0.14; 0.02	DECR
NIGER	0.63	0.28; −0.12; 0.47	YES	2.21	−0.14; 0.39; 1.96	INCR
NIGERIA	1.13	0.16; 0.39; 0.58	INCR	1.02	1.38; −0.65; 0.29	YES
MACEDONIA	0.74	1.83; −1.16; 0.07	YES	0.74	1.26; −0.10; −0.42	DECR
NORWAY	0.77	−0.19; −0.61; 1.57	YES	2.13	6.02; −10.80; 6.91	YES
OMAN	3.70	0.39; 0.12; 3.19	YES	9.80	−16.87; 39.41; −12.74	INV
PAKISTAN	1.22	−0.61; 1.07; 0.76	INV	1.19	0.55; −0.11; 0.75	YES
PALESTINE	0.96	−0.18; −0.23; 1.37	YES	1.06	−0.21; 0.18; 1.09	INCR
PANAMA	0.96	0.16; 0.56; 0.24	INV	0.79	1.22; −0.16; −0.27	DECR
PAPAU NEW G.	0.49	0.35; −1.96; 2.10	YES	0.88	−0.39; 0.04; 1.23	INCR
PARAGUAY	0.59	−1.52; 1.90; 0.21	INV	1.20	−3.20;3.06; 1.34	INV
PERU	0.89	8.30; −2.47; −4.94	DECR	0.53	3.98; −4.72; 1.27	YES
PHILIPPINES	1.15	0.89; −0.08; 0.34	YES	1.54	0.07; 2.84; −1.37	INV
POLAND	0.92	2.32; −1.89; 0.49	YES	1.31	1.71; −1.63; 1.23	YES
POLYNESIA	0.66	0.22; 0.20; 0.24	YES	0.21	−1.05; 1.09; 0.17	INV
PORTUGAL	1.56	−1.34; −8.29; 11.19	YES	3.89	1.13; −4.00; 6.76	YES
QATAR	0.80	−0.84; −1.99; 3.63	YES	1.03	0.62; 0.61; −0.20	INV
ROMANIA	0.88	0.90; 0.06; −0.08	DECR	0.95	1.23; −0.48; 0.20	YES
RUSSIA	1.07	1.16; −1.00; 0.91	YES	0.87	0.83; −5.77; 5.81	YES
RWANDA	1.80	3.20; 2.20; −3.60	DECR	0.14	3.93; −2.75; −1.04	YES
SAO TOME	1.44	0.44; 0.64; 0.36	INV	2.67	2.25; −3.45; 3.87	YES
SAN MARINO	5.10	0.28; 1.14;3.68	INCR	0.26	−0.05; 2.32; −2.01	INV
SAUDI ARABIA	0.90	−1.70; 2.94; −0.34	INV	0.98	−1.05; 0.54; 1.49	INCR
SENEGAL	0.72	−0.19; 1.48; −0.57	INV	1.59	0.73; 0.23; 0.63	YES
SERBIA	1.62	−0.40; 0.47; 1.55	INCR	0.82	2.02; −0.94; −0.26	YES
SEYCHELLES	0.48	0.30; 0.51; −0.33	INV	0.54	0.38; −0.19; 0.35	YES
SIERRA LEONE	2.23	−2.93; −0.80; 5.96	INCR	1.37	0.95; −1.25; 1.67	YES
SINGAPORE	1.33	1.15; 0.51; −0.33	DECR	2.83	1.61; −2.44; 3.66	YES
SLOVAK	0.99	−2.67; 1.90; 1.76	INV	0.74	0.97; −0.73; 0.50	YES

Country Name		**First wave**			**Second wave**	
	R_0	R_j**'s**	**U-shape**	R_0	R_j**'s**	**U-shape**
SLOVENIA	0.75	1.56; −0.71; −0.10	DECR	0.64	1.47; −0.47; −0.36	YES
SOMALIA	1.18	−0.16; 1.51; −0.17	INV	0.29	0.86; 0.57; −1.14	DECR
SOUTH AFRICA	0.87	0.22; 0.73; −0.08	INV	1.49	0.20; −0.04; 1.33	YES
SOUTH SUDAN	0.58	0.10; 0.16; 0.32	INCR	1.72	0.63; −0.63; 1.72	YES
SPAIN	0.38	−0.18; 0.27; 0.29	INCR	0.51	1.21; −0.86; 0.16	YES
SRI LANKA	2.13	2.73; −0.75; 0.15	YES	0.79	0.42; 1.00; −0.63	INV
ST KITTS NEVIS	2.00	0.00; 1.00; 1.00	INCR	1.07	0.25; 0.18; 0.64	YES
ST LUCIA	1.13	−0.53; −0.04; 1.70	INCR	1.00	1.00; −1.00; 1.00	YES
ST VINCENT	0.04	−0.29; 0.24; 0.10	INV	0.69	−0.24; 0.35; 0.58	INCR
SUDAN	0.36	−1.46; 2.34; −0.52	INV	2.00	0.00; 2.00; 0.00	INV
SURINAME	10.34	2.70; 18.77; −11.13	INV	1.63	2.95; −1.25; −0.07	YES
SWEDEN	0.56	0.58; −1.20; 1.18	YES	1.21	0.67; −0.91; 1.45	YES
SWITZERLAND	1.21	1.25; 0.13; −0.17	DECR	0.28	0.89; 1.18; −1.79	INV
SYRIA	1.43	1.39; 4.13; −4.09	INV	0.18	0.31; −0.68; 0.55	YES
TAIWAN	1.88	−0.13; 1.38; 0.63	INV	0.66	−5.21; 13.83; −7.96	INV
TAJIKISTAN	1.02	0.71; −0.60; 0.91	YES	1.49	1.83; −0.17; −0.17	YES
TANZANIA	0.91	−1.50; 0.18; 2.23	INCR	1.89	3.42; 14.26; −15.79	INV
THAILAND	0.69	0.42; 0.07; 0.20	YES	2.71	−1.77; −0.75; 5.23	INCR
TIMOR LESTE	5.00	1.00; 0.00; 4.00	YES	1.33	0.00; 1.00; 0.33	INV
TOGO	0.08	6.05; −6.18; 0.21	YES	1.14	0.18; 0.09; 0.87	YES
TRINIDAD	0.32	−0.26; 1.46; −0.88	INV	0.55	0.26; 0.03; 0.26	YES
TUNISIA	1.53	0.77; −0.04; 0.80	YES	2.77	−3.21; −2.41; 8.39	INCR
TURKEY	1.15	−1.50; −1.13; 3.78	INCR	2.21	19.82; −47.90; 30.29	YES
UAE	0.97	2.07; −1.11; 0.01	YES	1.15	1.25; −0.64; 0.54	YES
UGANDA	0.95	0.87; −0.37; 0.45	YES	0.64	0.44; −0.06; 0.26	YES
UKRAINE	0.96	1.35; −1.04; 0.65	YES	0.30	3.10; 1.07; −1.73	DECR
UK	0.76	−0.02; −0.76; 1.54	YES	1.03	0.43; 0.82; −0.22	INV
USA	8.42	31.42; −99.18; 76.18	YES	0.49	3.32; −0.38; −2.45	DECR
URUGUAY	0.63	0.71; 0.31; −0.39	DECR	1.03	−0.23; 0.35; 0.91	INCR
UZBEKISTAN	0.95	0.04; 0.10; 0.81	INCR	0.90	−0.03; −0.39; 1.32	YES
VENEZUELA	1.54	1.65; 2.95; −3.06	INV	0.82	1.09; −2.53; 2.26	YES
VIETNAM	3.29	−0.84; −0.39; 4.52	YES	1.43	0.76; −0.11; 0.78	YES
VIRGIN ISLANDS	0.51	0.01; −0.06; 0.56	YES	0.33	0.44; −0.22; 0.11	YES
WEST GAZA	1.00	−1.00; −2.00; 4.00	YES	0.98	0.59; −0.11; 0.50	YES
YEMEN	0.70	−0.34; 0.17; 0.86	INCR	1.50	1.00; 0.00; 0.50	YES
ZAMBIA	0.75	0.25; −0.13; 0.63	YES	1.12	1.11; −0.44; 0.45	YES
ZIMBABWE	1.44	0.24; 0.60; 0.60	INCR	1.62	1.08; −1.12; 1.66	YES

Appendix B

Appendix

Table B.1: Parameters for first and second waves for 195 countries.

Country Name	**First wave**			**Second wave**			**CHE/GDP**
	$\mathbf{R_o}$	**Exp. slope**	**A-slope**	$\mathbf{R_o}$	**Exp. slope**	**A-slope**	
AFGHANISTAN	1.78	0.1070	−0.025	0.79	0.0017	−0.097	9.40
ALGERIA	2.19	0.1594	−0.040	0.86	0.0316	−0.100	6.22
ARUBA	-	-	-	-	0.0172	−0.112	-
ANDORRA	-	−0.0313	−0.121	-	−0.0067	−0.155	6.71
ANGOLA	2.06	0.0100	−0.095	1.13	−0.0135	−0.057	2.55
ANTIGUA	4.23	-	-	3.30	0.0051	−0.177	5.23
ALBANIA	1.61	0.0091	−0.138	0.99	0.0058	−0.085	5.26
ARGENTINA	2.06	0.1485	−0.060	1.19	0.0427	−0.240	9.62
ARMENIA	1.51	0.0809	−0.050	0.80	0.0570	−0.090	10.03
AUSTRALIA	2.45	0.1832	−0.054	1.11	0.0037	−0.136	9.28
AUSTRIA	2.93	0.2825	−0.053	1.05	0.0034	−0.053	10.33
AZERBAIJAN	2.11	0.1422	−0.071	0.63	0.0676	−0.130	3.51
BAHAMAS	6.33	-	-	1.48	−0.0250	−0.077	6.25
BAHRAIN	1.81	0.1884	−0.079	1.24	0.0012	−0.053	4.13
BANGLADESH	3.67	0.0799	−0.033	0.92	−0.0086	−0.046	2.34
BARBADOS	4.63	-	-	1.99	0.0378	−0.109	6.56
BELARUS	3.15	0.0043	−0.060	1.02	0.0159	−0.026	5.64
BELGIUM	8.28	0.1963	−0.047	0.88	−0.0182	−0.063	10.32
BELIZE	3.74	-	-	1.34	−0.0004	−0.140	5.69
BENIN	2.16	0.0226	−0.133	1.55	0.0020	−0.125	2.49
BHUTAN	2.10	0.0021	−0.118	2.49	0.0126	−0.099	3.06
BOLIVIA	1.46	0.0647	−0.045	1.45	0.0152	−0.087	6.30
BOSNIA	1.70	0.0088	−0.110	0.97	−0.0118	−0.106	8.90
BOTSWANA	3.76	-	-	1.43	0.0030	−0.186	5.85
BRAZIL	3.10	0.0389	−0.048	0.92	0.0092	−0.188	9.51
BRUNEI	5.00	−0.0165	−0.120	3.66	-	-	2.41
BULGARIA	1.97	0.0178	−0.087	0.78	0.0049	−0.110	7.35
BURKINA FASO	2.44	−0.0227	−0.123	1.18	0.0360	−0.058	5.63
BURUNDI	2.80	-	-	1.69	0.0226	−0.063	7.74
CABO VERDE	1.54	0.0247	−0.091	1.71	−0.0064	−0.110	5.36
CAMBODIA	5.55	−0.0129	−0.129	3.12	0.0010	−0.158	6.03

Country Name	First wave			Second wave			CHE/GDP
	R_o	Exp. slope	A-slope	R_o	Exp. slope	A-slope	
CAMEROON	2.56	0.0338	−0.123	1.64	0.0085	−0.207	-
CANADA	2.95	0.2432	−0.029	1.05	0.0153	−0.047	10.79
CAR	2.45	−0.0130	−0.096	4.99	-	-	10.99
CHAD	2.43	−0.0108	−0.114	1.44	0.0222	−0.050	4.10
CHILE	2.42	0.1906	−0.034	1.16	0.0586	−0.090	9.14
CHINA	2.05	−0.0602	−0.088	1.07	0.0137	−0.068	5.35
COLUMBIA	1.86	0.0384	−0.040	0.99	0.0061	−0.126	7.64
COMOROS	1.93	−0.0094	−0.153	1.58	0.0397	−0.076	4.59
CONGO DEM.	1.48	0.0384	−0.052	1.10	0.0252	−0.089	3.30
CONGO REP.	2.39	0.0294	−0.152	1.43	0.0064	−0.118	2.14
COSTA RICA	1.51	0.0142	−0.110	1.08	−0.0022	−0.209	7.56
COTE D'VOIRE	1.47	0.0309	−0.080	1.35	0.0253	−0.078	4.19
CROTIA	3.95	−0.0042	−0.069	0.72	−0.0115	−0.106	6.83
CUBA	2.23	0.0706	−0.063	1.30	0.0517	−0.040	11.19
CURACAO	-	-	-	-	−0.0060	−0.074	-
CYPRUS	2.21	−0.0056	−0.131	1.30	0.0273	−0.089	6.77
CZECH	2.40	0.2570	−0.067	1.22	0.0474	−0.197	7.65
DENMARK	1.60	−0.0024	−0.087	0.90	0.0092	−0.048	10.07
DJIBOUTI	2.73	0.0144	−0.094	1.47	−0.0045	−0.169	2.32
DOMINICAN	2.09	0.0309	−0.088	1.10	0.0151	−0.081	5.73
DOMINICA	-	-	-	-	-	-	6.59
ECUADOR	2.22	0.0157	−0.140	1.18	−0.0045	−0.175	8.14
EGYPT	1.69	0.0527	−0.042	1.33	0.0243	−0.023	4.95
EL SALVADOR	1.58	0.0783	−0.052	1.29	0.0535	−0.113	7.11
EQUATORIAL G.	10.0	0.0454	−0.190	2.41	0.0142	−0.177	3.00
ERITREA	2.57	0.0083	−0.216	0.74	0.0222	−0.146	4.09
ESTONIA	2.10	−0.0254	−0.116	1.03	0.0279	−0.099	6.69
ESWATINI	2.08	0.0317	−0.071	1.34	0.0412	−0.034	6.54
ETHIOPIA	2.42	0.1259	−0.054	1.11	−0.0041	−0.136	3.30
FIJI	-	-		-	-	-	3.42
FINLAND	1.66	−0.0030	−0.093	1.04	−0.0010	−0.119	9.04
FRANCE	2.68	0.2898	−0.110	1.00	−0.0096	−0.081	11.26
GABON	1.83	0.0404	−0.077	1.44	0.0187	−0.143	2.75
GAMBIA	3.21	−0.0026	−0.094	2.29	0.0145	−0.099	3.09
GEORGIA	2.19	0.2536	−0.136	0.76	0.0293	−0.057	7.11
GERMANY	2.84	0.2624	−0.050	0.98	0.0050	−0.195	11.43
GHANA	1.85	0.0463	−0.099	1.09	0.0118	−0.117	3.45
GREECE	1.72	−0.0189	−0.091	1.05	−0.0111	−0.069	7.72
GRENADA	5.78	-	-	1.08	0.0106	−0.167	4.46
GUADELOUPE	-	−0.0131	−0.130	-	−0.0084	−0.137	-
GUATEMALA	1.67	0.0880	−0.044	1.08	0.1109	−0.197	5.71
GUIANA FRENCH	-	0.0391	−0.102	-	0.0238	−0.124	-
GUINEA	1.50	0.0097	−0.111	1.36	−0.0108	−0.126	3.93
GUINEA BISSAU	3.56	0.0230	−0.145	4.66	-	-	7.00
GUYANA	2.49	0.0005	−0.152	1.54	−0.0021	−0.163	5.94

Country Name	First wave			Second wave			CHE/GDP
	R_o	Exp. slope	A-slope	R_o	Exp. slope	A-slope	
HAITI	2.32	0.0565	−0.047	1.66	0.0217	−0.082	7.69
HONDURAS	1.96	0.0532	−0.086	1.59	0.0016	−0.141	7.05
HONG KONG	-	−0.0003	−0.060	-	0.0285	−0.041	-
HUNGARY	2.25	0.0018	−0.093	0.77	−0.0081	−0.088	6.70
ICELAND	2.89	−0.0261	−0.056	1.86	−0.0174	−0.079	8.47
INDIA	2.43	0.0331	−0.050	0.91	−0.0151	−0.048	3.54
INDONESIA	2.04	0.0391	−0.071	1.07	0.0127	−0.051	2.87
IRAN	3.61	0.2641	−0.063	1.00	0.0438	−0.140	8.66
IRAQ	1.81	0.1184	−0.084	0.96	0.0410	−0.150	-
IRELAND	2.63	−0.0021	−0.058	1.45	0.0188	−0.057	6.93
ISRAEL	2.86	−0.0047	−0.049	1.33	0.0339	−0.037	7.52
ITALY	2.99	0.2475	−0.040	1.06	−0.0057	−0.072	8.67
JAMAICA	2.43	−0.0031	−0.089	1.22	0.0034	−0.174	6.06
JAPAN	1.91	0.0872	−0.055	1.21	0.0260	−0.052	10.95
JORDAN	2.16	−0.0006	−0.155	0.93	−0.0138	−0.053	7.79
KAZAKHSTAN	2.85	0.0856	−0.064	1.05	0.0933	−0.210	2.92
KENYA	1.57	0.0413	−0.067	1.26	−0.0237	−0.310	5.17
KOREA REP.	6.06	0.1664	−0.076	0.90	0.0585	−0.090	7.56
KOSOVO	1.90	-	-	0.82	-	-	-
KUWAIT	2.25	0.0687	−0.031	1.27	−0.0094	−0.038	5.00
KYRGYZSTAN	2.27	0.0671	−0.091	0.86	0.0271	−0.200	6.53
LAO PDR	-	-	-	-	-	-	2.25
LATVIA	2.32	−0.0179	−0.087	1.10	0.0224	−0.136	6.19
LEBANON	1.91	0.2286	−0.112	1.27	0.0757	−0.180	8.35
LESOTHO	1.99	0.0053	−0.206	1.36	0.0398	−0.087	9.28
LIBERIA	1.76	0.0151	−0.114	3.08	0.0046	−0.159	6.74
LIBYA	3.12	0.0493	−0.047	1.09	−0.0059	−0.099	-
LITHUANIA	1.63	0.0394	−0.096	0.98	0.0554	−0.230	6.57
LUXEMBOURG	1.99	−0.0401	−0.061	0.83	−0.0174	−0.105	5.29
MACAO	-	−0.0019	−0.190	-	-	-	-
MADAGASCAR	2.48	0.0377	−0.057	1.54	0.0060	−0.211	4.79
MALAWI	3.55	0.0478	−0.088	1.66	0.0583	−0.087	9.33
MALAYSIA	2.86	0.1042	−0.101	1.15	0.0794	−0.260	3.76
MALDIVES	1.96	0.0031	−0.154	1.41	0.0007	−0.116	9.41
MALI	1.61	0.0158	−0.100	0.97	0.0148	−0.115	-
MALTA	4.46	0.0712	−0.114	1.29	0.0536	−0.330	8.96
MAURITANIA	1.66	−0.0033	−0.055	0.82	0.0362	−0.037	4.58
MAURITIUS	5.40	−0.0209	−0.120	9.32	−0.0032	−0.143	5.83
MAYOTTE	-	0.0129	−0.103	-	0.0065	−0.154	-
MEXICO	2.03	0.1759	−0.100	0.98	0.0117	−0.109	5.37
MOLDOVA	2.03	0.0324	−0.086	0.83	−0.0037	−0.127	6.60
MONACO	5.48	−0.0044	−0.147	1.66	0.0134	−0.136	1.60
MONGOLIA	3.12	0.0116	−0.204	1.98	0.0195	−0.127	3.79
MONTENEGRO	8.16	−0.0114	−0.171	1.07	0.0040	−0.085	8.42
MOROCCO	2.05	0.1161	−0.114	0.84	−0.0159	−0.065	5.31

Country Name	First wave			Second wave			CHE/GDP
	R_o	Exp. slope	A-slope	R_o	Exp. slope	A-slope	
MOZAMBIQUE	2.14	0.0260	−0.109	1.59	0.0152	−0.068	8.17
MYANMAR	2.70	−0.0028	−0.113	0.83	−0.0137	−0.050	4.79
NAMIBIA	2.10	0.0315	−0.049	1.03	0.0315	−0.039	7.95
NEPAL	2.28	0.2070	−0.035	0.91	−0.0264	−0.065	5.84
NETHERLANDS	2.40	0.2485	−0.043	0.92	0.0002	−0.074	9.97
NEW ZEALAND	5.63	−0.0426	−0.087	1.89	0.0140	−0.099	9.21
NICARAGUA	5.76	-	-	1.39	-	-	8.56
NIGER	2.58	−0.0231	−0.083	0.96	0.0390	−0.048	7.33
NIGERIA	1.91	0.0502	−0.046	1.06	0.0333	−0.047	3.89
MACEDONIA	1.84	0.0858	−0.092	0.87	0.0528	−0.230	6.58
NORWAY	2.40	0.2716	−0.055	1.14	0.0052	−0.145	10.05
OMAN	1.73	0.0972	−0.092	1.13	0.0936	−0.130	4.13
PAKISTAN	1.90	0.1301	−0.060	1.02	0.0113	−0.047	3.20
PALESTINE	-	−0.0053	−0.202	-	0.0063	−0.050	-
PANAMA	2.08	0.1443	−0.063	1.13	0.1195	−0.070	7.27
PAPAU NEW G.	1.95	−0.0081	−0.115	2.45	-	-	2.37
PARAGUAY	2.22	0.0196	−0.147	0.97	0.0032	−0.168	6.65
PERU	2.35	0.0915	−0.010	1.26	−0.0077	−0.111	5.24
PHILIPPINES	2.29	0.1627	−0.082	0.91	0.1772	−0.174	4.40
POLAND	2.17	0.1562	−0.079	0.99	0.0094	−0.072	6.33
POLYNESIA	-	-	-	-	-	−0.075	-
PORTUGAL	2.92	0.0301	−0.140	1.15	0.0431	−0.190	9.41
QATAR	2.61	0.0694	−0.070	1.16	−0.0019	−0.094	2.49
ROMANIA	2.26	0.0218	−0.056	0.91	−0.0072	−0.121	5.56
RUSSIA	2.41	0.0775	−0.020	1.00	0.0046	−0.037	5.32
RWANDA	2.03	0.0615	−0.146	1.26	0.0382	−0.064	7.54
SAO TOME	3.09	−0.0218	−0.153	3.33	0.0162	−0.127	6.27
SAN MARINO	5.88	−0.0157	−0.137	1.14	−0.0028	−0.154	7.14
SAUDI ARABIA	2.31	0.0607	−0.060	0.90	−0.0138	−0.029	6.36
SENEGAL	2.02	0.0351	−0.047	1.24	0.0387	−0.047	3.98
SERBIA	2.13	0.0042	−0.053	0.79	0.0123	−0.038	8.54
SEYCHELLES	2.68	-	-	1.94	0.0313	−0.134	5.11
SIERRA LEONE	1.50	0.0143	−0.107	1.52	0.0291	−0.063	16.06
SINGAPORE	2.06	0.0551	−0.030	1.52	0.0641	−0.080	4.46
SLOVAK	1.74	−0.0286	−0.123	0.92	0.0028	−0.193	6.69
SLOVENIA	1.78	−0.0345	−0.079	1.08	−0.0004	−0.263	8.30
SOLOMON ISL.	-	-	-	-	-	-	- 4.47
SOMALIA	1.95	−0.0085	−0.091	2.55	-	-	-
SOUTH AFRICA	2.54	0.257	−0.110	1.15	0.0303	−0.039	8.25
SOUTH SUDAN	2.99	0.0007	−0.152	1.59	0.0095	−0.133	6.40
SPAIN	3.85	0.3350	−0.035	1.16	0.0029	−0.080	8.98
SRI LANKA	4.14	0.0144	−0.159	1.04	0.1347	−0.160	3.76
ST KITTS NEVIS	-	-	-	-	-	-	5.31
ST LUCIA	1.34	-	-	2.86	0.0157	−0.082	4.40
ST VINCENT	5.86	-	-	2.17	0.0407	−0.080	4.47

Country Name	First wave			Second wave			CHE/GDP
	R_o	Exp. slope	A-slope	R_o	Exp. slope	A-slope	
SUDAN	1.97	0.0193	−0.094	0.24	0.0407	−0.039	4.51
SURINAME	1.41	0.0214	−0.061	1.27	0.0379	−0.042	7.97
SWEDEN	2.10	0.2572	−0.106	1.05	0.0123	−0.162	10.90
SWITZERLAND	2.86	0.2388	−0.044	0.95	−0.0082	−0.041	11.88
SYRIA	2.80	0.0311	−0.030	1.06	0.0086	−0.041	-
TAIWAN	3.42	0.0036	−0.084	1.84	0.0053	−0.123	-
TAJIKISTAN	1.68	0.0418	−0.066	0.54	−0.0016	−0.131	7.24
TANZANIA	5.00	0.1205	−0.125	18.4	-	-	3.63
THAILAND	3.42	−0.0201	−0.055	1.63	0.0496	−0.100	3.79
TIMOR LESTE	-	-	-	-	-	-	4.33
TOGO	2.09	0.0093	−0.084	1.41	0.0083	−0.112	6.17
TRINIDAD	5.35	−0.0025	−0.139	1.30	−0.0102	−0.113	6.93
TUNISIA	2.64	−0.0122	−0.084	1.15	0.0053	−0.117	7.29
TURKEY	4.32	0.0120	−0.040	0.81	0.0078	−0.030	4.12
UAE	2.33	0.0484	−0.080	1.22	0.0085	−0.055	4.23
UGANDA	2.18	-	-	0.88	0.0047	−0.154	6.53
UKRAINE	2.16	0.0325	−0.130	0.89	−0.0032	−0.093	7.72
UK	2.89	0.2223	−0.037	1.25	0.0106	−0.035	10.00
USA	3.85	0.2882	−0.030	0.99	0.0121	−0.060	16.89
URUGUAY	2.76	−0.0228	−0.086	1.15	0.0389	−0.039	9.20
UZBEKISTAN	1.82	0.1231	−0.088	0.71	0.0238	−0.170	5.29
VENEZUELA	2.57	0.0389	−0.073	0.94	0.0002	−0.134	3.56
VIETNAM	3.59	−0.0166	−0.158	1.94	−0.0040	−0.158	5.92
WEST GAZA	3.73	-	-	0.87	-	-	-
YEMEN	1.57	0.0049	−0.164	2.84	0.0006	−0.150	-
ZAMBIA	2.80	0.0265	−0.134	1.73	0.0372	−0.046	4.93
ZIMBABWE	1.98	0.0367	−0.087	1.40	0.0438	−0.045	4.73

Table B.2: Africa countries data.

S/N	Country Name	SDG index	EPI	TBI	HDI	SI	LE	Temperature
1	Algeria	65.8	57.18	-	21	72.22	78.0	18.79
2	Angola	49.3	37.44	3.51	-	58.33	61.0	21.55
3	Benin	51.5	38.17	0.55	3.2	43.52	61.0	27.55
4	Botswana	61.6	51.70	-	20	62.04	65.0	21.50
5	Bukina Faso	53.5	42.83	0.47	0.6	22.22	63.0	28.29
6	Burundi	50.3	27.43	1.07	-	20.37	67.0	20.80
7	Cabo Verde	64.1	56.94	0.46	19	45.37	73.0	23.30
8	Cameroon	51.6	40.81	1.79	1.4	35.19	62.0	24.60
9	CAR	36.7	36.42	5.40	2.9	18.52	54.0	24.90
10	Chad	38.7	45.34	1.42	-	22.22	58.0	26.55
11	Comoros	47.6	44.24	0.35	4.2	-	66.0	25.55

S/N	Country Name	SDG index	EPI	TBI	HDI	SI	LE	Temperature
12	DRC	41.6	30.41	3.20	0.1	-	61.0	24.00
13	Congo	48.7	42.39	3.73	-	47.22	61.0	24.55
14	Ivory Coast	55.6	45.25	1.37	21	25.93	61.0	26.35
15	Djibouti	49.7	40.04	-	40	37.96	65.0	28.00
16	Egypt	63.8	61.21	-	3.6	54.63	74.0	22.10
17	Equatorial	43.4	60.40	1.81	-	-	66.0	24.55
18	Eritrea	43.3	39.34	0.86	0.1	86.11	66.0	25.50
19	Eswatini	52.4	-	3.63	5.0	68.52	59.0	21.40
20	Ethiopia	53.2	44.78	1.40	8.4	50.93	68.0	22.20
21	Gabon	59.4	45.05	5.21	20	50.93	69.0	25.05
22	Gambia	51.9	42.42	1.58	4.3	44.44	66.0	27.50
23	Ghana	61.2	49.66	1.44	2.5	50.93	68.0	27.20
24	Guinea	49.4	46.62	1.76	3.1	52.78	63.0	25.70
25	Guinea Bissau	45.5	44.67	3.61	1.3	-	63.0	26.75
26	Kenya	56.6	47.25	2.67	0.9	50.93	69.0	24.75
27	Lesotho	50.9	33.78	6.54	6.8	65.74	53.0	11.85
28	Liberia	48.0	41.62	3.08	-	40.74	65.0	25.30
29	Libya	-	49.79	-	34	76.85	77.0	21.80
30	Madagascar	45.6	33.73	2.33	4.9	37.04	67.0	22.65
31	Malawi	52.3	49.21	1.46	1.9	48.15	63.0	21.90
32	Mali	51.7	43.71	0.52	2.5	48.15	62.0	28.25
33	Mauritania	51.3	39.24	0.89	14	29.63	65.0	27.65
34	Mauritius	66.2	56.63	0.12	3.0	93.52	76.0	22.40
35	Morocco	64.4	63.47	-	19	76.85	73.0	18.10
36	Mozambique	51.4	46.37	3.61	5.1	55.56	56.0	23.80
37	Namibia	57.1	58.46	4.86	1.4	42.59	65.0	20.95
38	Niger	50.3	35.74	0.84	7.3	34.26	59.0	27.15
39	Nigeria	47.1	54.76	2.19	0.1	45.37	60.0	26.80
40	Rwanda	57.9	43.68	0.57	1.8	71.30	65.0	18.85
41	Sao Tome	61.8	54.01	1.14	-	-	66.0	23.75
42	Senegal	57.0	49.52	1.17	12	13.89	63.0	27.85
43	Seychelles	-	66.02	0.16	-	65.74	76.0	27.15
44	Sierra Leone	49.7	42.54	2.95	6.6	48.15	60.0	23.82
45	Somalia	40.1	-	2.58	-	47.22	54.0	27.05
46	South Africa	60.4	44.73	6.15	1.0	48.15	65.0	18.75
47	South Sudan	29.2	-	2.27	-	72.22	-	-
48	Sudan	47.4	51.49	0.67	5.4	24.07	67.0	26.90
49	Tanzania	55.9	50.83	2.37	16	8.33	64.0	22.35
50	Togo	52.7	41.78	0.37	1.1	67.59	67.0	27.15
51	Tunisia	66.1	62.35	-	5.4	62.04	76.0	16.30
52	Uganda	54.9	44.28	2.00	0.5	53.70	68.0	22.80
53	Zambia	53.0	50.97	3.33	20	36.11	54.0	21.40
54	Zimbabwe	54.8	43.41	1.99	16	53.70	62.0	21.00

Table B.3: Comparison between the predicted and actual values.

Day	Deaths		Tests	
	Actual/%gain	**Predicted**	**Actual**	**Predicted**
1	49/43%	86	269886	120499
2	46/66%	136	251301	126261
3	50/42%	86	248354	178764
4	153/-32%	116	248910	264431
5	26/68%	82	96177	192823
6	11/72%	39	30345	75319
7	53/45%	97	247760	328075
8	78/42%	136	222942	124271
9	43/65%	124	206626	118126
10	52/70%	171	207651	237759
11	150/14%	175	214336	124387
12	39/69%	126	86361	229623
13	27/36%	42	23804	78612
14	81/-12%	72	223293	268357
15	85/51%	172	199948	360891
16	63/62%	168	191917	135525
17	63/74%	246	196259	199318
18	130/11%	146	210495	168640
19	49/71%	167	90639	160399
20	32/56%	59	25699	249580
21	69/62%	181	240612	294387
22	66/64%	185	217585	298154
23	80/57%	188	214258	150142
24	76/61%	194	231306	150122
25	109/50%	219	259073	64680
26	54/74%	209	114369	301318
27	46/41%	78	32368	45122
28	95/52%	198	299121	343681
29	108/63%	292	276013	274313
30	104/37%	165	279376	325882
31	88/68%	275	301465	284502
32	178/44%	320	322468	262879
33	89/63%	238	140298	312368
34	85/50%	169	40313	154521
35	146/67%	447	355160	390516
36	262/30%	376	321373	254298
37	163/60%	410	330328	419636
38	162/51%	329	357368	445595
39	298/38%	484	388884	217528
40	137/33%	206	165764	242920
41	116/64%	318	47485	223540
42	257/40%	430	430644	264886
43	523/-63%	320	387569	256737
44	244/55%	548	379590	484870

Bibliography

[1] D. Bernoulli. Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir. *Mém. Math. Phys. Acad. Roy. Sci. Paris*, pages 1–45, 1760.

[2] H. Nishiura. Time variations in the transmissibility of pandemic influenza in prussia, germany, from 1918–19. *Theoretical Biology and Medical Modelling*, 4(1), 2007. doi: 10.1186/1742-4682-4-20.

[3] A. M. Fry, K. Hancock, M. Patel, M. Gladden, S. Doshi, D. M. Blau, D. Sugerman, V. Veguilla, X. Lu, H. Noland, Y. Bai, A. Maroufi, A. Kao, P. Kriner, K. Lopez, M. Ginsberg, S. Jain, S. J. Olsen, and J. M. Katz. The first cases of 2009 pandemic influenza a (h1n1) virus infection in the united states: a serologic investigation demonstrating early transmission. *Influenza and Other Respiratory Viruses*, 6(3), 2012. doi: 10.1111/j.1750-2659.2012.00339.x.

[4] M. Worobey, T. D. Watts, R. A. McKay, M. A. Suchard, T. Granade, D. E. Teuwen, B. A. Koblin, W. Heneine, P. Lemey, and H. W. Jaffe. 1970s and 'patient 0' hiv-1 genomes illuminate early hiv/aids history in north america. *Nature*, 539(7627):98–101, 2016. doi: 10.1038/nature19827.

[5] S. E. F. Yong, D. E. Anderson, W. E. Wei, J. Pang, W. N. Chia, C. W. Tan, Y. L. Teoh, P. Rajendram, M. P. H. S. Toh, C. Poh, V. T. J. Koh, J. Lum, N-A. Suhaimi, P. Y. Chia, M. I-C. Chen, S. Vasoo, B. Ong, Y. S. Leo, L. Wang, and V. J. M. Lee. Connecting clusters of covid-19: an epidemiological and serological investigation. *The Lancet Infectious Diseases*, 20(7):809–815, 2020. doi: 10.1016/s1473-3099(20)30273-5.

[6] D. C. Adam, P. Wu, J. Y. Wong, E. H. Y. Lau, T. K. Tsang, S. Cauchemez, G. M. Leung, and B. J. Cowling. Clustering and superspreading potential of sars-cov-2 infections in hong kong. *Nature Medicine*, 26(11):1714–1719, 2020. doi: 10.1038/s41591-020-1092-0.

[7] M.R. Desjardins, A. Hohl, and E.M. Delmelle. Rapid surveillance of covid-19 in the united states using a prospective space-time scan statistic: Detecting and evaluating emerging clusters. *Applied Geography*, 118:102202–102202, 2020. doi: 10.1016/j.apgeog.2020.102202.

[8] R. Pung, C. J. Chiew, B. E. Young, S. Chin, M. I-C. Chen, H. E. Clapham, A. R. Cook, S. Maurer-Stroh, M. P. H. S. Toh, C. Poh, M. Low, J. Lum, V. T. J. Koh, T. M. Mak, L. Cui, R. V. T. P. Lin, D. Heng, Y-S. Leo, D. C. Lye, V. J. M. Lee, K q. Kam, S. Kalimuddin, S. Y. Tan, J. Loh, K. C. Thoon, S. Vasoo, W. X. Khong, N-A. Suhaimi,

S. J. H. Chan, E. Zhang, O. Oh, A. Ty, C. Tow, Y. X. Chua, W. L. Chaw, Y. Ng, F. Abdul-Rahman, S. Sahib, Z. Zhao, C. Tang, C. Low, E. H. Goh, G. Lim, Y. Hou, I. Roshan, J. Tan, K. Foo, K. Nandar, L. Kurupatham, and P. P. Chan. Investigation of three clusters of covid-19 in singapore: implications for surveillance and response measures. *The Lancet*, 395(10229):1039–1046, 2020. doi: 10.1016/s0140-6736(20)30528-6.

[9] T. Ganyani, C. Kremer, D. Chen, A. Torneri, C. Faes, J. Wallinga, and N. Hens. Estimating the generation interval for coronavirus disease (covid-19) based on symptom onset data, march 2020. *Eurosurveillance*, 25(17), 2020. doi: 10.2807/1560-7917.es.2020.25.17.2000257.

[10] C. Kremer, T. Ganyani, D. Chen, A. Torneri, C. Faes, J. Wallinga, and N. Hens. Authors' response: Estimating the generation interval for covid-19 based on symptom onset data. *Eurosurveillance*, 25(29), 2020. doi: 10.2807/1560-7917.es.2020.25.29.2001269.

[11] Q-L. Jing, M-J. Liu, Z-B. Zhang, L-Q. Fang, J. Yuan, A-R. Zhang, N. E. Dean, L. Luo, M-M. Ma, I. Longini, E. Kenah, Y. Lu, Y. Ma, N. Jalali, Z-C. Yang, and Y. Yang. Household secondary attack rate of covid-19 and associated determinants in guangzhou, china: a retrospective cohort study. *The Lancet Infectious Diseases*, 20(10):1141–1150, 2020. doi: 10.1016/s1473-3099(20)30471-0.

[12] M. M. Böhmer, U. Buchholz, V. M. Corman, M. H., K. Katz, D. V. Marosevic, S. Böhm, T. Woudenberg, N. Ackermann, R. Konrad, U. Eberle, B. Treis, A. Dangel, K. Bengs, V. Fingerle, A. Berger, S. Hörmansdorfer, S. Ippisch, B. Wicklein, A. Grahl, K. Pörtner, N. Muller, N. Zeitlmann, T. S. Boender, W. Cai, A. Reich, M. an der Heiden, U. Rexroth, O. Hamouda, J. Schneider, T. Veith, B. Mühlemann, R. Wölfel, M. Antwerpen, M. Walter, U. Protzer, B. Liebl, W. Haas, A. Sing, C. Drosten, and A. Zapf. Investigation of a covid-19 outbreak in germany resulting from a single travel-associated primary case: a case series. *The Lancet Infectious Diseases*, 20(8):920–928, 2020. doi: 10.1016/s1473-3099(20)30314-5.

[13] A. Guttmann, L. Ouchchane, X. Li, I. Perthus, J. Gaudart, J. Demongeot, and J-Y. Boire. Performance map of a cluster detection test using extended power. *International Journal of Health Geographics*, 12(1):47–47, 2013. doi: 10.1186/1476-072x-12-47.

[14] F. Shams, A. Abbas, W. Khan, U. S. Khan, and R. Nawaz. A death, infection, and recovery (dir) model to forecast the covid-19 spread. *Computer Methods and Programs in Biomedicine Update*, 2:100047–100047, 2022. doi: 10.1016/j.cmpbup.2021.100047.

[15] T-C. Chan and C-C. King. Surveillance and epidemiology of infectious diseases using spatial and temporal lustering methods. *Infectious Disease Informatics and Biosurveillance*, pages 207–234, 2010. doi: 10.1007/978-1-4419-6892-0_10.

[16] S. Hisada, T. Murayama, K. Tsubouchi, S. Fujita, S. Yada, S. Wakamiya, and E. Aramaki. Surveillance of early stage covid-19 clusters using search query logs and mobile device-based location information. *Scientific Reports*, 10(1), 2020. doi: 10.1038/s41598-020-75771-6.

[17] L. Han, P. Shen, J. Yan, Y. Huang, X. Ba, W. Lin, H. Wang, Y. Huang, K. Qin, Y. Wang, Z. Chen, and S. Tu. Exploring the clinical characteristics of covid-19 clusters identified using factor analysis of mixed data-based cluster analysis. *Frontiers in Medicine*, 8, 2021. doi: 10.3389/fmed.2021.644724.

[18] A. Ladoy, O. Opota, P-N. Carron, I. Guessous, S. Vuilleumier, S. Joost, and G. Greub. Size and duration of covid-19 clusters go along with a high sars-cov-2 viral load: A spatio-temporal investigation in vaud state, switzerland. *Science of The Total Environment*, 787:147483–147483, 2021. doi: 10.1016/j.scitotenv.2021.147483.

[19] A. Tariq, Y. Lee, K. Roosa, S. Blumberg, P. Yan, S. Ma, and G. Chowell. Real-time monitoring the transmission potential of covid-19 in singapore, march 2020. *BMC Medicine*, 18(1), 2020. doi: 10.1186/s12916-020-01615-9.

[20] World health organization. https://www.who.int/emergencies/diseases/novel-coronavirus-2019, 2020. Accessed: 14 April 2020.

[21] N. S. AlTakarli. China's response to the covid-19 outbreak: A model for epidemic preparedness and management. *Dubai Medical Journal*, 3(2):44–49, 2020. doi: 10.1159/000508448.

[22] J. Mackolil and B. Mahanthesh. Logistic growth and sir modelling of coronavirus disease (covid-19) outbreak in india: Models based on real-time data. *Mathematical Modelling of Engineering Problems*, 7(3):345–350, 2020. doi: 10.18280/mmep.070303.

[23] M.A. Khanday and F. Zargar. Mathematical analysis on the dynamics of covid-19 in india using sir epidemic model. *Mapana Journal of Sciences*, 19(3), 2020. doi: 10.12723/mjs.54.1.

[24] R. Yaghoby. *A new nonlinear mathematical model for forecasting the waves of the COVID-19 epidemic*. Research Square Platform LLC, 2021. doi: 10.21203/rs.3.rs-243403/v1.

[25] E. Kaxiras and G. Neofotistos. Multiple epidemic wave model of the covid-19 pandemic: Modeling study. *Journal of Medical Internet Research*, 22(7), 2020. doi: 10.2196/20912.

[26] A. Meiksin. Using the seir model to constrain the role of contaminated fomites in spreading an epidemic: An application to covid-19 in the uk. *Mathematical Biosciences and Engineering*, 19(4):3564–3590, 2022. doi: 10.3934/mbe.2022164.

[27] M. Cámara, M. Miravete, and E. Navarro. *An epidemic model for economical impact predicting and spatiotemporal spreading of COVID-19*. Cold Spring Harbor Laboratory, 2020. doi: 10.1101/2020.09.02.20186551.

[28] M. Paunovic. Mathematical modeling and statistical analysis in post - covid-19 tourism recovery. *Tourism Challenges Amid COVID-19*, 2021. doi: 10.52370/tisc21334mp.

[29] Y. Duan. Simulations of the covid-19 epidemic in nigeria using sir model. *Journal of Physics: Conference Series*, 1893(1):12016–12016, 2021. doi: 10.1088/1742-6596/1893/1/012016.

[30] V. Padmapriya and M. Kaliyappan. Fuzzy fractional mathematical model of covid-19 epidemic. *Journal of Intelligent & Fuzzy Systems*, 42(4):3299–3321, 2022. doi: 10.3233/jifs-211173.

[31] K. Kozioł, R. Stanisławski, and G. Bialic. Fractional-order sir epidemic model for transmission prediction of covid-19 disease. *Applied Sciences*, 10(23):8316–8316, 2020. doi: 10.3390/app10238316.

[32] K. Deforche. *An age-structured epidemiological model of the Belgian COVID-19 epidemic*. Cold Spring Harbor Laboratory, 2020. doi: 10.1101/2020.04.23.20077115.

[33] F. Balabdaoui and D. Mohr. Age-stratified discrete compartment model of the covid-19 epidemic with application to switzerland. *Scientific Reports*, 10(1), 2020. doi: 10.1038/s41598-020-77420-4.

[34] C. M. Verrelli and F. D. Rossa. Two-age-structured covid-19 epidemic model: Estimation of virulence parameters to interpret effects of national and regional feedback interventions and vaccination. *Mathematics*, 9(19):2414–2414, 2021. doi: 10.3390/math9192414.

[35] H. Liu and L. Wang. Focus on the mental health of pediatric medical workers in china after the covid-19 epidemic. *Frontiers in Psychology*, 12, 2021. ISSN 1664-1078. doi: 10.3389/fpsyg.2021.657814.

[36] X. Li and H. Lyu. Epidemic risk perception, perceived stress, and mental health during covid-19 pandemic: A moderated mediating model. *Frontiers in Psychology*, 11, 2021. doi: 10.3389/fpsyg.2020.563741.

[37] F. A. Rihan, H. J. Alsakaji, and C. Rajivganthi. Stochastic sirc epidemic model with time-delay for covid-19. *Advances in Difference Equations*, 2020(1), 2020. doi: 10.1186/s13662-020-02964-8.

[38] H. Kang, Y. Lou, G. Chen, S. Chu, and X. Fu. Epidemic spreading and global stability of a new sis model with delay on heterogeneous networks. *Journal of Biological Systems*, 23(4):1550029–1550029, 2015. doi: 10.1142/s0218339015500291.

[39] L. Dell'Anna. Solvable delay model for epidemic spreading: the case of covid-19 in italy. *Scientific Reports*, 10(1), 2020. doi: 10.1038/s41598-020-72529-y.

[40] G. Bianconi. Epidemic spreading. *Oxford Scholarship Online*, 2018. doi: 10.1093/oso/9780198753919.003.0013.

[41] F. Nian, X. Guo, and J. Li. A new spreading model in the environment of epidemic-related online rumors. *Modern Physics Letters B*, 36(4), 2021. doi: 10.1142/s0217984921505692.

[42] B. Zhao. Epidemic-related sites in covid-19 media reports. *Clinical Research and Clinical Trials*, 2(3):1–10, 2020. doi: 10.31579/2693-4779/005.

[43] B. B. Süt. Epigenetic regulation mechanisms in viral infections: A special focus on covid-19. *Biotechnology to Combat COVID-19*, 2022. doi: 10.5772/intechopen.98866.

[44] J. C. Jayaraj, D. Mahalingam, and G. Ganesan. Covid-19: When the misinformation is spreading faster than the virus. *Journal of Thoracic Disease and Cardiothoracic Surgery*, 1(1):1–1, 2020. doi: 10.31579/2693-2156/005.

[45] S. Zihady. Multiple variable social distancing model in reducing covid 19 spreading: A long term social distancing model in developing nations like bangladesh. *Public Administration Research*, 10(2):66–66, 2021. doi: 10.5539/par.v10n2p66.

[46] S. Gounane, Y. Barkouch, A. Atlas, M. Bendahmane, F. Karami, and D. Meskine. An adaptive social distancing sir model for covid-19 disease spreading and forecasting. *Epidemiologic Methods*, 10, 2021. doi: 10.1515/em-2020-0044.

[47] G. Webb. A covid-19 epidemic model predicting the effectiveness of vaccination. *Mathematics in Applied Sciences and Engineering*, pages 1–15, 2021. doi: 10.5206/mase/13889.

[48] A. Temerev, L. Rozanova, O. Keiser, and J. Estill. Geospatial model of covid-19 spreading and vaccination with event gillespie algorithm. *Nonlinear Dynamics*, 2022. doi: 10.1007/s11071-021-07186-5.

[49] R. T. Alqahtani and A. Yusuf. Development and analysis of a seir model for covid-19 epidemic with vaccination and non-singular kernel. *Fractals*, 2021. doi: 10.1142/s0218348x22400400.

[50] G. Webb. A covid-19 epidemic model predicting the effectiveness of vaccination in the us. *Infectious Disease Reports*, 13(3):654–667, 2021. doi: 10.3390/idr13030062.

[51] D. Hidayat and E. S. Nugraha. Epidemic model analysis of covid-19. *E3S Web of Conferences*, 328:6002–6002, 2021. doi: 10.1051/e3sconf/202132806002.

[52] G. Gavalas. *A Lattice Model for COVID-19 Epidemic*. MDPI AG, 2020. doi: 10.20944/preprints202009.0367.v1.

[53] S. O. Ilyin. *COVID-19 spreading: a recursive model*, volume 7. Cold Spring Harbor Laboratory, 2021. doi: 10.2196/21468.

[54] D. Below. Comparative analysis of the spread of the covid 19 epidemic in berlin and new york city based on a computational model. *Public Health Open Access*, 5(1), 2021. doi: 10.23880/phoa-16000176.

[55] C. Pizzuti, A. Socievole, B. Prasse, and P. V. Mieghem. Network-based prediction of covid-19 epidemic spreading in italy. *Applied Network Science*, 5(1), 2020. doi: 10.1007/s41109-020-00333-8.

[56] C. Zuo, A. Wang, F. Zhu, Z. Meng, and X. Zhao. A new coupled awareness-epidemic spreading model with neighbor behavior on multiplex networks. *Complexity*, 2021:1–14, 2021. doi: 10.1155/2021/6680135.

[57] A. Arenas, J. Gómez-Gardeñes, C. Granell, and D. Soriano-Paños. Epidemic spreading: Tailored models for covid-19. *Europhysics News*, 51(5):38–40, 2020. doi: 10.1051/epn/2020507.

[58] A. Bianconi, A. Marcelli, G. Campi, and A. Perali. Ostwald growth rate in controlled covid-19 epidemic spreading as in arrested growth in quantum complex matter. *Condensed Matter*, 5(2):23–23, 2020. doi: 10.3390/condmat5020023.

[59] D. Prodanov. Analytical parameter estimation of the sir epidemic model. applications to the covid-19 pandemic. *Entropy*, 23(1):59–59, 2020. doi: 10.3390/e23010059.

[60] V. H. Moreau. Using the weibull distribution to model covid-19 epidemic data. *Model Assisted Statistics and Applications*, 16(1):5–14, 2021. doi: 10.3233/mas-210510.

[61] Z. Xu, H. Zhang, and Z. Huang. A continuous markov-chain model for the simulation of covid-19 epidemic dynamics. *Biology*, 11(2):190–190, 2022. doi: 10.3390/biology11020190.

[62] Y. Tsori and R. Granek. Epidemiological model for the inhomogeneous spatial spreading of covid-19 and other diseases. *PLOS ONE*, 16(2), 2021. doi: 10.1371/journal.pone.0246056.

[63] R-C. Tsaur, C-L. Chiu, and Y-Y. Huang. Fuzzy portfolio selection in covid-19 spreading period using fuzzy goal programming model. *Mathematics*, 9(8):835–835, 2021. doi: 10.3390/math9080835.

[64] G. Zhang and X. Liu. Prediction and control of covid-19 spreading based on a hybrid intelligent model. *PLOS ONE*, 16(2), 2021. doi: 10.1371/journal.pone.0246360.

[65] V. Zakharov, Y. Balykina, O. Petrosian, and H. Gao. Cbrr model for predicting the dynamics of the covid-19 epidemic in real time. *Mathematics*, 8(10):1727–1727, 2020. doi: 10.3390/math8101727.

[66] C. Bereketoglu, N. Ozcan, T. Raika Kiran, and Mehmet Lutfi Yola. Forecasting model to predict the spreading of the covid-19 outbreak in turkey. *Journal of Intelligent Systems with Applications*, pages 95–102, 2021. doi: 10.54856/jiswa.202112165.

[67] F. T. Akyildiz and F. S. Alshammari. Complex mathematical sir model for spreading of covid-19 virus with mittag-leffler kernel. *Advances in Difference Equations*, 2021(1), 2021. doi: 10.1186/s13662-021-03470-1.

[68] B. Jamshidi, M. Rezaei, and K. Rezaei. A new model for epidemic spreading with a focus on covid-19. *Health Scope*, 9(3), 2020. ISSN 2251-8959. doi: 10.5812/jhealthscope.102837.

[69] A. K. Cherkashin. Hierarchical epidemic risk modeling of spreading new covid-19 coronavirus. *Issues of Risk Analysis*, 17(4):10–21, 2020. doi: 10.32686/1812-5220-2020-17-4-10-21.

[70] J. D'Alembert. Opuscules mathématiques. *David: Paris, France*, 2:35–38, 1761.

[71] K. Dietz and J.A.P. Heesterbeek. Daniel bernoulli's epidemiological model revisited. *Mathematical Biosciences*, 180(1):1–21, 2002. doi: 10.1016/s0025-5564(02)00122-0.

[72] R. Ross. An application of the theory of probabilities to the study of a priori pathometry.—part i. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 92(638):204–230, 1916. doi: 10.1098/rspa.1916.0007.

[73] A. G. M'Kendrick. Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, 44:98–130, 1925. doi: 10.1017/S0013091500034428.

[74] P. H. Leslie. Some further notes on the use of matrices in population mathematics. *Biometrika*, 35(3):213–213, 1948. doi: 10.2307/2332342.

[75] P. H. Leslie. On the use of matrices in certain population mathematics. *Biometrika*, 33(3):183–212, 1945. doi: 10.1093/biomet/33.3.183.

[76] L. A. Demetrius. Boltzmann, darwin and directionality theory. *Physics Reports*, 530 (1):1–85, 2013. doi: 10.1016/j.physrep.2013.04.001.

[77] C. J. Rhodes and L. A. Demetrius. Evolutionary entropy determines invasion success in emergent epidemics. *PLoS ONE*, 5(9), 2010. doi: 10.1371/journal.pone.0012951.

[78] M. B. Usher. A matrix model for forest management. *Biometrics*, 25(2):309–309, 1969. doi: 10.2307/2528791.

[79] Frédéric Mortier, Vivien Rossi, Gilles Guillot, Sylvie Gourlet-Fleury, and Nicolas Picard. Population dynamics of species-rich ecosystems: the mixture of matrix population models approach. *Methods in Ecology and Evolution*, 4(4):316–326, 2012. doi: 10.1111/2041-210x.12019.

[80] M. Zetlaoui, N. Picard, and A. Bar-Hen. Asymptotic distribution of stage-grouped population models. *Mathematical Biosciences*, 200(1):76–89, 2006. doi: 10.1016/j.mbs.2005.12.003.

[81] M. Geurts, G. E. P. Box, and G. M. Jenkins. Time series analysis: Forecasting and control. *Journal of Marketing Research*, 14(2):269–269, 1977. doi: 10.2307/3150485.

[82] R. J. Hyndman, A. B. Koehler, R. D. Snyder, and S. Grose. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3):439–454, 2002. doi: 10.1016/s0169-2070(01)00110-8.

[83] P. Furtado. Epidemiology sir with regression, arima, and prophet in forecasting covid-19. *The 7th International conference on Time Series and Forecasting*, 2021. doi: 10.3390/engproc2021005052.

[84] W. D. Ray and M. B. Priestley. Spectral analysis and time series. vol. 1: Univariate series. *Journal of the Royal Statistical Society. Series A (General)*, 145(2):269–269, 1982. doi: 10.2307/2981548.

[85] S. N. Wood. *Generalized Additive Models*. Chapman and Hall/CRC, 2017. doi: 10.1201/9781315370279.

[86] M. Krzyśko and L. Waszak. Canonical correlation analysis for functional data. *Biometrical Letters*, 50(2):95–105, 2013. doi: 10.2478/bile-2013-0020.

[87] I. T. Jolliffe and J. Cadima. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202–20150202, 2016. doi: 10.1098/rsta.2015.0202.

[88] S. Lee, J. Kim, J. Hwang, E. Lee, K-J. Lee, J. Oh, J. Park, and T-Y. Heo. Clustering of time series water quality data using dynamic time warping: A case study from the bukhan river water quality monitoring network. *Water*, 12(9):2411–2411, 2020. doi: 10.3390/w12092411.

[89] A. S. Hadi, L. Kaufman, and P. J. Rousseeuw. Finding groups in data: An introduction to cluster analysis. *Technometrics*, 34(1):111–111, 1992. doi: 10.2307/1269576.

[90] M. A. Mahfouz. Efficient single-linkage hierarchical clustering based on partitioning. *2016 11th International Conference on Computer Engineering & Systems (ICCES)*, 2016. doi: 10.1109/icces.2016.7821986.

[91] F. M. Salem. Gated rnn: The long short-term memory (lstm) rnn. *Recurrent Neural Networks*, pages 71–82, 2021. doi: 10.1007/978-3-030-89929-5_4.

[92] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9 (8):1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735.

[93] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley. Deep learning for healthcare: review, opportunities and challenges. *Briefings in Bioinformatics*, 19(6):1236–1246, 2017. doi: 10.1093/bib/bbx044.

[94] D. Ravi, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G-Z. Yang. Deep learning for health informatics. *IEEE Journal of Biomedical and Health Informatics*, 21(1):4–21, 2017. doi: 10.1109/jbhi.2016.2636665.

[95] O. Triebe, H. Hewamalage, P. Pilyugina, N. Laptev, C. Bergmeir, and R. Rajagopal. Neuralprophet: Explainable forecasting at scale. *medRxiv [Preprint]*, 2021. doi: 10.48550/ARXIV.2111.15397.

[96] Wikipedia. https://www.wikipedia.org/wiki/, 2021. Accessed: 14 November 2021.

[97] J. O. Ramsay, G. Hooker, and S. Graves. Introduction to functional data analysis. *Functional Data Analysis with R and MATLAB*, pages 1–19, 2009. doi: 10.1007/978-0-387-98185-7_1.

[98] J. Strandberg, S. S. de Luna, and J. Mateu. A comparison of spatiotemporal and functional kriging approaches. *Geostatistical Functional Data Analysis*, pages 375–402, 2021. doi: 10.1002/9781119387916.ch15.

[99] J. Jacques and C. Preda. Functional data clustering: a survey. *Advances in Data Analysis and Classification*, 8(3):231–255, 2013. doi: 10.1007/s11634-013-0158-y.

[100] F. Ieva, A. M. Paganoni, D. Pigoli, and V. Vitelli. Multivariate functional clustering for the morphological analysis of electrocardiograph curves. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62(3):401–418, 2012. doi: 10.1111/j.1467-9876.2012.01062.x.

[101] J. Demongeot, Y. Flet-Berliac, and H. Seligmann. Temperature decreases spread parameters of the new covid-19 case dynamics. *Biology*, 9(5):94–94, 2020. doi: 10.3390/biology9050094.

[102] H. Seligmann, S. Iggui, M. Rachdi, Nicolas Vuillerme, and Jacques Demongeot. Inverted covariate effects for first versus mutated second wave covid-19: High temperature spread biased for young. *Biology*, 9(8):226–226, 2020. doi: 10.3390/biology9080226.

[103] H. Seligmann, N. Vuillerme, and J. Demongeot. Unpredictable, counter-intuitive geo-climatic and demographic correlations of covid-19 spread rates. *Biology*, 10(7):623–623, 2021. doi: 10.3390/biology10070623.

[104] Y. Pan, D. Zhang, P. Yang, L. L. M. Poon, and Q. Wang. Viral load of sars-cov-2 in clinical samples. *The Lancet Infectious Diseases*, 20(4):411–412, 2020. doi: 10.1016/s1473-3099(20)30113-4.

[105] M. Delbrück. Statistical fluctuations in autocatalytic reactions. *The Journal of Chemical Physics*, 8(1):120–124, 1940. doi: 10.1063/1.1750549.

[106] J. Rubio. Stability analysis of the modified levenberg–marquardt algorithm for the artificial neural network training. *IEEE Transactions on Neural Networks and Learning Systems*, 32(8):3510–3524, 2021. doi: 10.1109/tnnls.2020.3015200.

[107] J. Rubio, E. Lughofer, J. Pieper, P. Cruz, D. I. Martinez, Genaro Ochoa, Marco Antonio Islas, and Enrique Garcia. Adapting h-infinity controller for the desired reference tracking of the sphere position in the maglev process. *Information Sciences*, 569:669–686, 2021. doi: 10.1016/j.ins.2021.05.018.

[108] L. Böttcher, O. Woolley-Meza, E. Goles, D. Helbing, and H. J. Herrmann. Connectivity disruption sparks explosive epidemic spreading. *Physical Review E*, 93(4), 2016. doi: 10.1103/physreve.93.042315.

[109] S-N. Chow and J. K. Hale. Methods of bifurcation theory. *Grundlehren der mathematischen Wissenschaften*, 1982. doi: 10.1007/978-1-4613-8159-4.

[110] P. lacoude, covid-19: Le début de la fin? https://www.contrepoints.org/2020/07/22/376624-covid-19-lx10-debut-dx10-la-fin-1, 2020. Accessed: 14 January 2020.

[111] R. Wölfel, V. M. Corman, W. Guggemos, M. Seilmaier, S. Zange, M. A. Müller, D. Niemeyer, T. C. Jones, P. Vollmar, C. Rothe, M. Hoelscher, T. Bleicker, S. Brünink, J. Schneider, R. Ehmann, K. Zwirglmaier, C. Drosten, and C. Wendtner. Virological assessment of hospitalized patients with covid-2019. *Nature*, 581(7809):465–469, 2020. doi: 10.1038/s41586-020-2196-x.

[112] W-D. Liu, S-Y. Chang, J-T. Wang, M-J. Tsai, C-C. Hung, C-L. Hsu, and S-C. Chang. Prolonged virus shedding even after seroconversion in a patient with covid-19. *Journal of Infection*, 81(2):318–356, 2020. doi: 10.1016/j.jinf.2020.03.063.

[113] L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, and C. Fraser. Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing. *Science*, 368(6491), 2020. doi: 10.1126/science.abb6936.

[114] H-Y. Cheng, S-W. Jian, D-P. Liu, T-C. Ng, W-T. Huang, and H-H. Lin. Contact tracing assessment of covid-19 transmission dynamics in taiwan and risk at different exposure periods before and after symptom onset. *JAMA Internal Medicine*, 180(9): 1156–1156, 2020. doi: 10.1001/jamainternmed.2020.2020.

[115] Xi. He, E. H. Y. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, X. Mo, Y. Chen, B. Liao, W. Chen, F. Hu, Q. Zhang, M. Zhong, Y. Wu, L. Zhao, F. Zhang, B. J. Cowling, F. Li, and G. M. Leung. Temporal dynamics in viral shedding and transmissibility of covid-19. *Nature Medicine*, 26(5):672–675, 2020. doi: 10.1038/s41591-020-0869-5.

[116] D. L. Chao, M. E. Halloran, V. J. Obenchain, and I. M. Longini. Flute, a publicly available stochastic influenza epidemic simulation model. *PLoS Computational Biology*, 6(1), 2010. doi: 10.1371/journal.pcbi.1000656.

[117] Y. Itoh, S. Shichinohe, M. Nakayama, M. Igarashi, A. Ishii, H. Ishigaki, H. Ishida, N. Kitagawa, T. Sasamura, M. Shiohara, M. Doi, H. Tsuchiya, S. Nakamura, M. Okamatsu, Y. Sakoda, H. Kida, and K. Ogasawara. Emergence of h7n9 influenza a virus resistant to neuraminidase inhibitors in nonhuman primates. *Antimicrobial Agents and Chemotherapy*, 59(8):4962–4973, 2015. doi: 10.1128/aac.00793-15.

[118] J. Demongeot and C. Taramasco. Evolution of social networks: the example of obesity. *Biogerontology*, 15(6):611–626, 2014. doi: 10.1007/s10522-014-9542-z.

[119] Worldometers. https://www.worldometers.info/coronavirus/, 2021. Accessed: 14 November 2021.

[120] J. Demongeot and L. A. Demetrius. Complexity and stability in biological systems. *International Journal of Bifurcation and Chaos*, 25(7):1540013–1540013, 2015. doi: 10.1142/s0218127415400131.

[121] N. L. Garcia. Birth and death processes as projections of higher-dimensional poisson processes. *Advances in Applied Probability*, 27(4):911–930, 1995. doi: 10.1017/s000186780004773x.

[122] R. Breban, R. Vardavas, and S. Blower. Theory versus data: How to calculate r0? *PLoS ONE*, 2(3), 2007. doi: 10.1371/journal.pone.0000282.

[123] J. T. Wu, K. Leung, M. Bushman, N. Kishore, R. Niehus, P. M. de Salazar, B. J. Cowling, M. Lipsitch, and G. M. Leung. Estimating clinical severity of covid-19 from the transmission dynamics in wuhan, china. *Nature Medicine*, 26(4):506–510, 2020. doi: 10.1038/s41591-020-0822-7.

[124] J. Demongeot, O. Hansen, and C. Taramasco. Discrete dynamics of contagious social diseases: Example of obesity. *Virulence*, 7(2):129–140, 2015. doi: 10.1080/21505594.2015.1082708.

[125] J. Demongeot, M. Jelassi, and C. Taramasco. From susceptibility to frailty in social networks: The case of obesity. *Mathematical Population Studies*, 24(4):219–245, 2017. doi: 10.1080/08898480.2017.1348718.

[126] Renkulab. https://renkulab.shinyapps.io/COVID-19-Epidemic-Forecasting/wf89fdf90/, 2021. Accessed: 14 November 2021.

[127] J. Scire, S. Nadeau, T. Vaughan, G. Brupbacher, S. Fuchs, J. Sommer, K. N. Koch, R. Misteli, L. Mundorff, T. Gtz, T. Eichenberger, C. Quinto, M. Savic, A. Meienberg, T. Burkard, M. Mayr, C. A. Meier, A. Widmer, R. Kuehl, A. Egli, H. H. Hirsch, S. Bassetti, C. H. Nickel, K. S. Rentsch, W. Kbler, R. Bingisser, M. Battegay, S. Tschudin-Sutter, and T. Stadler. Reproductive number of the covid-19 epidemic in switzerland with a focus on the cantons of basel-stadt and basel-landschaft. *Swiss Medical Weekly*, 2020. doi: 10.4414/smw.2020.20271.

[128] L. A. Demetrius. Demographic parameters and natural selection. *Proceedings of the National Academy of Sciences*, 71(12):4645–4647, 1974. doi: 10.1073/pnas.71.12.4645.

[129] L. A. Demetrius. Statistical mechanics and population biology. *Journal of Statistical Physics*, 30(3):709–753, 1983. ISSN 0022-4715. doi: 10.1007/bf01009685.

[130] J. Demongeot and L. A. Demetrius. Sélection naturelle et dérive démographique. Étude empirique de la france de 1860 à 1965. *Population (French Edition)*, 44(1):109–109, 1989. doi: 10.2307/1533336.

[131] J. Demongeot. Biological boundaries and biological age. *Acta Biotheoretica*, 57(4): 397–418, 2009. doi: 10.1007/s10441-009-9087-8.

[132] J. Gaudart, M. Ghassani, J. Mintsa, M. Rachdi, J. Waku, and J. Demongeot. Demography and diffusion in epidemics: Malaria and black death spread. *Acta Biotheoretica*, 58(2):277–305, 2010. doi: 10.1007/s10441-010-9103-z.

[133] J. Demongeot, O. Hansen, H. Hessami, A. S. Jannot, J. Mintsa, M. Rachdi, and C. Taramasco. Random modelling of contagious diseases. *Acta Biotheoretica*, 61(1):141–172, 2013. doi: 10.1007/s10441-013-9176-6.

[134] A. D. Ventsel' and M. I. Freidlin. On small random perturbations of dynamical systems. *Russian Mathematical Surveys*, 25(1):1–55, 1970. doi: 10.1070/rm1970v025n01abeh001254.

[135] M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain markov process expectations for large time, ii. *Communications on Pure and Applied Mathematics*, 28 (2):279–301, 1975. doi: 10.1002/cpa.3160280206.

[136] M. I. Freidlin and A. D. Wentzell. Random perturbations. *Grundlehren der mathematischen Wissenschaften*, pages 15–43, 1984. doi: 10.1007/978-1-4684-0176-9_2.

[137] J. Demongeot, Q. Griette, and P. Magal. Si epidemic model applied to covid-19 data in mainland china. *Royal Society Open Science*, 7(12):201878–201878, 2020. doi: 10.1098/rsos.201878.

[138] J. Gaudart, J. Landier, L. Huiart, E. Legendre, L. Lehot, M. K. Bendiane, L. Chiche, A. Petitjean, E. Mosnier, F. Kirakoya-Samadoulougou, J. Demongeot, R. Piarroux, and S. Rebaudet. Factors associated with the spatial heterogeneity of the first wave of covid-19 in france: a nationwide geo-epidemiological study. *The Lancet Public Health*, 6(4), 2021. doi: 10.1016/s2468-2667(21)00006-2.

[139] J. Demongeot and H. Seligmann. Sars-cov-2 and mirna-like inhibition power. *Medical Hypotheses*, 144:110245–110245, 2020. doi: 10.1016/j.mehy.2020.110245.

[140] Z. Liu, P. Magal, O. Seydi, and G. Webb. Understanding unreported cases in the covid-19 epidemic outbreak in wuhan, china, and the importance of major public health interventions. *Biology*, 9(3):50–50, 2020. doi: 10.3390/biology9030050.

[141] Q. Griette and P. Magal. Clarifying predictions for covid-19 from testing data: The example of new york state. *Infectious Disease Modelling*, 6:273–283, 2021. doi: 10.1016/j.idm.2020.12.011.

[142] J. Demongeot, H. Hazgui, H. B. Amor, and J. Waku. Stability, complexity and robustness in population dynamics. *Acta Biotheoretica*, 62(3):243–284, 2014. doi: 10.1007/s10441-014-9229-5.

[143] Q. Griette, J. Demongeot, and P. Magal. A robust phenomenological approach to investigate covid-19 data for france. *Mathematics in Applied Sciences and Engineering*, pages 1–12, 2021. doi: 10.5206/mase/14031.

[144] J. Demongeot, M. Jelassi, H. Hazgui, S. B. Miled, N. B. B. Saoud, and C. Taramasco. Biological networks entropies: Examples in neural memory networks, genetic regulation networks and social epidemic networks. *Entropy*, 20(1):36–36, 2018. doi: 10.3390/e20010036.

[145] J. Demongeot, Q. Griette, P. Magal, and G. Webb. Modeling vaccine efficacy for covid-19 outbreak in new york city. *Biology*, 11(3):345–345, 2022. doi: 10.3390/biology11030345.

[146] Q. Griette, J. Demongeot, and P. Magal. What can we learn from covid-19 data by using epidemic models with unidentified infectious cases? *Mathematical Biosciences and Engineering*, 19(1):537–594, 2021. doi: 10.3934/mbe.2022025.

[147] P. Magal, O. Seydi, G. Webb, and Y. Wu. A model of vaccination for dengue in the philippines 2016–2018. *Frontiers in Applied Mathematics and Statistics*, 7, 2021. doi: 10.3389/fams.2021.760259.

[148] M. Rachdi, J. Waku, H. Hazgui, and J. Demongeot. Entropy as a robustness marker in genetic regulatory networks. *Entropy*, 22(3):260–260, 2020. doi: 10.3390/e22030260.

[149] R.B. Ogunrinde, U.K. Nwajeri, S.E. Fadugba, R.R. Ogunrinde, and K.I. Oshinubi. Dynamic model of covid-19 and citizens reaction using fractional derivative. *Alexandria Engineering Journal*, 60(2):2001–2012, 2021. doi: 10.1016/j.aej.2020.09.016.

[150] Cameroon situation reports. https://reliefweb.int/report/cameroon/, 2022. Accessed: 10 February 2022.

[151] L. N. Nkamba and T. T. Manga. Modelling and prediction of the spread of covid-19 in cameroon and assessing the governmental measures (march–september 2020). *COVID*, 1(3):622–644, 2021. doi: 10.3390/covid1030052.

[152] A. Rahman and M.A. Kuddus. Modelling the transmission dynamics of covid-19 in six high burden countries. *BioMed Res. Int.*, 9(5089184), 2021. doi: 10.37473/fic/10.1101/2020.04.22.20075192.

[153] A. Ducrot, P. Magal, T. Nguyen, and G. F. Webb. Identifying the number of unreported cases in sir epidemic models. *Mathematical Medicine and Biology: A Journal of the IMA*, 37(2):243–261, 2019. doi: 10.1093/imammb/dqz013.

[154] C-J. Yu, Z-X. Wang, Y. Xu, M-X. Hu, K. Chen, and G. Qin. Assessment of basic reproductive number for covid-19 at global level. *Medicine*, 100(18), 2021. doi: 10.1097/md.0000000000025837.

[155] D. Tabutin and B. Schoumaker. The demography of sub-saharan africa in the 21st century. transformations since 2000, outlook to 2050. *Population*, 75(2):169–169, 2020. doi: 10.3917/popu.2002.0169.

[156] Knoema. https://knoema.com/atlas/Cameroon/topics/Demographics/Fertility/Fertility-rates, 2022. Accessed: 14 January 2021.

[157] Populationpyramid. https://www.populationpyramid.net, 2022. Accessed: 14 January 2022.

[158] Eurospe. https://abstracts.eurospe.org/hrp/0086/hrp0086p1-p749, 2022. Accessed: 14 January 2022.

[159] INS. *Enquête démographique et de Santé Cameroun 2004*. ORC Macro: Calverton, 2005.

[160] World bank. https://data.worldbank.org/, 2021. Accessed: 14 November 2021.

[161] Ministère de la Santé du Cameroun. *Rapport de situation COVID-19 n° 84*. Ministère de la Santé du Cameroun: Yaoundé, Cameroun, 2021.

[162] Santé publique france. https://www.santepubliquefrance.fr/dossiers/coronavirus-covid-19/coronavirus-437chiffres-cles-et-evolution-de-la-covid-19-en-france-et-dans-le-m 2022. Accessed: 14 January 2022.

[163] Epiet. https://www.ecdc.europa.eu/en/health-protection-surveillance-centre-epiet/, 2022. Accessed: 14 January 2022.

[164] Health population surveillance centre. https://www.hpsc.ie/az/respiratory/coronavirus/novelcoronavirus443/surveillance/epidemiologyofcovid-19inirelandweeklyreports/, 2022. Accessed: 14 January 2022.

[165] H. Caswell. A general formula for the sensitivity of population growth rate to changes in life history parameters. *Theoretical Population Biology*, 14(2):215–230, 1978. doi: 10.1016/0040-5809(78)90025-4.

[166] M. Ochs, J. R. Nyengaard, A. Jung, L. Knudsen, M. Voigt, T. Wahlers, J. Richter, and H. J. G. Gundersen. The number of alveoli in the human lung. *American Journal of Respiratory and Critical Care Medicine*, 169(1):120–124, 2004. doi: 10.1164/rccm.200308-1107oc.

[167] J. L. Wright and A. M. Churg. Macro and micro structure of the lung. *Clinical Respiratory Medicine*, pages 1–6, 2012. doi: 10.1016/b978-1-4557-0792-8.00001-5.

[168] P. Colombo, S. Cagnani, F. Sonvico, P. Santi, P. Russo, and G. Colombo. Biological in vitro models for absorption by nonoral routes. *Comprehensive Medicinal Chemistry II*, pages 279–299, 2007. doi: 10.1016/b0-08-045044-x/00127-9.

[169] B. M. Wiebe and H. Laursen. Human lung volume, alveolar surface area, and capillary length. *Microscopy Research and Technique*, 32(3):255–262, 1995. doi: 10.1002/jemt.1070320308.

[170] R. Sender, Y. M. Bar-On, S. Gleizer, B. Bernshtein, A. Flamholz, R. Phillips, and R. Milo. The total number and mass of sars-cov-2 virions. *Proceedings of the National Academy of Sciences*, 118(25), 2021. doi: 10.1073/pnas.2024815118.

[171] Nuccore. https://www.ncbinlmnihgov/nuccore, 2021. Accessed: 14 January 2021.

[172] Genecards. www.genecardsorg/cgi-bin/carddisppl?gene=HBB, 2021. Accessed: 14 January 2021.

[173] Mirbase. www.mirbaseorg/cgi-bin/mirna_entrypl?acc=MI0000093, 2021. Accessed: 14 January 2021.

[174] P. Xu, L. E. Palmer, C. Lechauve, G. Zhao, Y. Yao, J. Luan, A. Vourekas, H. Tan, J. Peng, J. D. Schuetz, Z. Mourelatos, G. Wu, M. J. Weiss, and V. R. Paralkar. Regulation of gene expression by mir-144/451 during mouse erythropoiesis. *Blood*, 133(23):2518–2528, 2019. doi: 10.1182/blood.2018854604.

[175] T. Wang, F. Wu, and D. Yu. mir-144/451 in hematopoiesis and beyond. *ExRNA*, 1(1), 2019. doi: 10.1186/s41544-019-0035-8.

[176] K. Lai, S. Jia, S. Yu, J. Luo, and Y. He. Genome-wide analysis of aberrantly expressed lncrnas and mirnas with associated co-expression and cerna networks in β-thalassemia and hereditary persistence of fetal hemoglobin. *Oncotarget*, 8(30):49931–49943, 2017. doi: 10.18632/oncotarget.18263.

[177] M. Rodríguez-López, S. Gonzalez, O. Hillson, E. Tunnacliffe, S. Codlin, V. A. Tallada, J. Bähler, and C. Rallis. The gata transcription factor gaf1 represses trnas, inhibits growth, and extends chronological lifespan downstream of fission yeast torc1. *Cell Reports*, 30(10):3240–3249, 2020. doi: 10.1016/j.celrep.2020.02.058.

[178] R. J. De Boer and A. S. Perelson. Quantifying t lymphocyte turnover. *Journal of Theoretical Biology*, 327:45–87, 2013. doi: 10.1016/j.jtbi.2012.12.025.

[179] J-P. Changeux, Z. Amoura, F. A. Rey, and M. Miyara. A nicotinic hypothesis for covid-19 with preventive and therapeutic implications. *Comptes Rendus. Biologies*, 343(1):33–39, 2020. doi: 10.5802/crbiol.8.

[180] H. Nishiura, K. Ito, A. Anzai, T. Kobayashi, C. Piantham, and A. J. Rodríguez-Morales. Relative reproduction number of sars-cov-2 omicron (b.1.1.529) compared with delta variant in south africa. *Journal of Clinical Medicine*, 11(1):30–30, 2021. doi: 10.3390/jcm11010030.

[181] J. Demongeot, E. Drouet, A. Elena, A. Moreira, Y. Rechoum, and S. Sené. Micro-rnas: viral genome and robustness of gene expression in the host. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1908): 4941–4965, 2009. doi: 10.1098/rsta.2009.0176.

[182] S. Bandiera, R. Matégot, M. Girard, J. Demongeot, and A. Henrion-Caude. Mitomirs delineating the intracellular localization of micrornas at mitochondria. *Free Radical Biology and Medicine*, 64:12–19, 2013. doi: 10.1016/j.freeradbiomed.2013.06.013.

[183] J. Demongeot, H. Hazgui, S. Bandiera, O. Cohen, and A. Henrion-Caude. Mitomirs, chloromirs and modelling of the microrna inhibition. *Acta Biotheoretica*, 61(3):367–383, 2013. doi: 10.1007/s10441-013-9190-8.

[184] S. Pfeffer, A. Sewer, M. Lagos-Quintana, R. Sheridan, C. Sander, F. A Grässer, L. F. van Dyk, C. K. Ho, S. Shuman, M. Chien, J. J. Russo, J. Ju, G. Randall, B. D. Lindenbach, C. M. Rice, V. Simon, D. D. Ho, M. Zavolan, and T. Tuschl. Identification of micrornas of the herpesvirus family. *Nature Methods*, 2(4):269–276, 2005. doi: 10.1038/nmeth746.

[185] R. P. Kincaid, J. M. Burke, and C. S. Sullivan. Rna virus microrna that mimics a b-cell oncomir. *Proceedings of the National Academy of Sciences*, 109(8):3077–3082, 2012. doi: 10.1073/pnas.1116107109.

[186] B. R. Cullen. Viruses and micrornas. *Nature Genetics*, 38, 2006. doi: 10.1038/ng1793.

[187] D. M. Dykxhoorn, L. D. Schlehuber, I. M. London, and J. Lieberman. Determinants of specific rna interference-mediated silencing of human β-globin alleles differing by a single nucleotide polymorphism. *Proceedings of the National Academy of Sciences*, 103 (15):5953–5958, 2006. doi: 10.1073/pnas.0601309103.

[188] D. M. Dykxhoorn and J. Lieberman. The silent revolution: Rna interference as basic biology, research tool, and therapeutic. *Annual Review of Medicine*, 56(1):401–423, 2005. doi: 10.1146/annurev.med.56.082103.104606.

[189] S. Yekta, I h. Shih, and D. P. Bartel. Microrna-directed cleavage of hoxb8 mrna. *Science*, 304(5670):594–596, 2004. doi: 10.1126/science.1097434.

[190] J. Brennecke, A. Stark, R. B. Russell, and S. M. Cohen. Principles of microrna–target recognition. *PLoS Biology*, 3(3), 2005. doi: 10.1371/journal.pbio.0030085.

[191] X. Li, Z. Fu, H. Liang, Y. Wang, X. Qi, M. Ding, X. Sun, Z. Zhou, Y. Huang, H. Gu, L. Li, X. Chen, D. Li, Q. Zhao, F. Liu, H. Wang, J. Wang, K. Zen, and C-Y. Zhang. H5n1 influenza virus-specific mirna-like small rna increases cytokine production and mouse mortality via targeting poly(rc)-binding protein 2. *Cell Research*, 28(2):157–171, 2018. doi: 10.1038/cr.2018.3.

[192] S-Y. L. Marzano, A. Neupane, and L. Domier. Transcriptional and small rna responses of the white mold fungus sclerotinia sclerotiorum to infection by a virulence-attenuating hypovirus. *Viruses*, 10(12):713–713, 2018. doi: 10.3390/v10120713.

[193] T-Y. Yu, M. Chen, and C-D. Wang. Annotation of mirnas in covid-19 coronavirus. *Journal of Electronic Science and Technology*, 19(1):100060–100060, 2021. doi: 10.1016/j.jnlest.2020.100060.

[194] R. Mirzaei, F. Mahdavi, F. Badrzadeh, S. R. Hosseini-Fard, M. H., A. S. Jeda, T. Mohammadi, M. Roshani, R. Yousefimashouf, H. Keyvani, M. Darvishmotevalli, M. Z. Sani, and S. Karampoor. The emerging role of micrornas in the severe acute respiratory syndrome coronavirus 2 (sars-cov-2) infection. *International Immunopharmacology*, 90: 107204–107204, 2021. doi: 10.1016/j.intimp.2020.107204.

[195] M. D. S. Demirci and A. Adan. Computational analysis of microrna-mediated interactions in sars-cov-2 infection. *PeerJ*, 8, 2020. doi: 10.7717/peerj.9369.

[196] F. Abedi, R. Rezaee, A. W. Hayes, S. Nasiripour, and G. Karimi. Micrornas and sars-cov-2 life cycle, pathogenesis, and mutations: biomarkers or therapeutic agents? *Cell Cycle*, 20(2):143–153, 2020. doi: 10.1080/15384101.2020.1867792.

[197] J. Gasparello, A. Finotti, and R. Gambari. Tackling the covid-19 "cytokine storm" with microrna mimics directly targeting the 3'utr of pro-inflammatory mrnas. *Medical Hypotheses*, 146:110415–110415, 2021. doi: 10.1016/j.mehy.2020.110415.

[198] N. Chauhan, M. Jaggi, S. C. Chauhan, and M. M. Yallapu. Covid-19: fighting the invisible enemy with micrornas. *Expert Review of Anti-infective Therapy*, 19(2):137–145, 2020. doi: 10.1080/14787210.2020.1812385.

[199] Q. Zhang, A. Honko, J. Zhou, H. Gong, S. N. Downs, J. H. Vasquez, R. H. Fang, W. Gao, A. Griffiths, and L. Zhang. Cellular nanosponges inhibit sars-cov-2 infectivity. *Nano Letters*, 20(7):5570–5574, 2020. doi: 10.1021/acs.nanolett.0c02278.

[200] S. Nersisyan, N. Engibaryan, A. Gorbonos, K. Kirdey, A. Makhonin, and A. Tonevitsky. Potential role of cellular mirnas in coronavirus-host interplay. *PeerJ*, 8, 2020. doi: 10.7717/peerj.9994.

[201] R. Satyam, T. Bhardwaj, S. Goel, N. K. Jha, S. K. Jha, P. Nand, J. Ruokolainen, M. A. Kamal, and K. K. Kesari. mirnas in sars-cov 2: A spoke in the wheel of pathogenesis. *Current Pharmaceutical Design*, 27(13):1628–1641, 2021. doi: 10.2174/1381612826999201001200529.

[202] A-A-K. Khan, R. U. Sany, S. Islam, and A. M. K. Islam. Epigenetic regulator mirna pattern differences among sars-cov, sars-cov-2, and sars-cov-2 world-wide isolates delineated the mystery behind the epic pathogenicity and distinct clinical characteristics of pandemic covid-19. *Frontiers in Genetics*, 11, 2020. doi: 10.3389/fgene.2020.00765.

[203] I. Bozgeyik. Therapeutic potential of mirnas targeting sars-cov-2 host cell receptor ace2. *Meta Gene*, 27:100831–100831, 2021. doi: 10.1016/j.mgene.2020.100831.

[204] B. C. Bernardo, P. Gregorevic, R. H. Ritchie, and J. R. McMullen. Generation of microrna-34 sponges and tough decoys for the heart: Developments and challenges. *Frontiers in Pharmacology*, 9, 2018. doi: 10.3389/fphar.2018.01090.

[205] G. M. Hahn. A formalism describing the kinetics of some mammalian cell populations. *Mathematical Biosciences*, 6:295–304, 1970. doi: 10.1016/0025-5564(70)90069-6.

[206] J. Demongeot, D. Benaouda, and C. Jézéquel. Dynamical confinement in neural networks and cell cycle. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 5(1): 167–173, 1995. doi: 10.1063/1.166064.

[207] J. Chu, A. Ducrot, P. Magal, and S. Ruan. Hopf bifurcation in a size-structured population dynamic model with random growth. *Journal of Differential Equations*, 247(3):956–1000, 2009. doi: 10.1016/j.jde.2009.04.003.

[208] B. Tang, F. Scarabel, N. L. Bragazzi, Z. McCarthy, M. Glazer, Y. Xiao, J. M. Heffernan, A. Asgary, N. H. Ogden, and J. Wu. De-escalation by reversing the escalation with a stronger synergistic package of contact tracing, quarantine, isolation and personal protection: Feasibility of preventing a covid-19 rebound in ontario, canada, as a case study. *Biology*, 9(5):100–100, 2020. doi: 10.3390/biology9050100.

[209] Q. Griette, P. Magal, and O. Seydi. Unreported cases for age dependent covid-19 outbreak in japan. *Biology*, 9(6):132–132, 2020. doi: 10.3390/biology9060132.

[210] M. Doumic. Analysis of a population model structured by the cells molecular content. *Mathematical Modelling of Natural Phenomena*, 2(3):121–152, 2007. doi: 10.1051/mmnp:2007006.

[211] R. Dilao and A. Lakmeche. Diffusion in the mckendrick-von foerster equation. *Proc. Dyn. Syst. Appl.*, 4:647–653, 2004. doi: 10.1051/mmnp:2006001.

[212] J-P. Caubet. Relativistic brownian motion. *Lecture Notes in Mathematics*, pages 113–142, 1975. doi: 10.1007/bfb0068583.

[213] M. Porte. *Passion des Formes. A René Thom.* ENS Editions: Paris, France, 1994.

[214] Our world in data covid-stringency. https://ourworldindata.org/grapher/covid-stringency-index-tab=table, 2021. Accessed: 14 November 2021.

[215] Drees. https://data.drees.solidarites-sante.gouv.fr/explore/dataset/covid-19-resultats-par-age-issus-des-appariements-entre-si-vic-si-dep-et-vac-si/analyze/, 2021. Accessed: 14 January 2021.

[216] N. Bacaër. *A Short History of Mathematical Population Dynamics*. Springer London, 2011. doi: 10.1007/978-0-85729-115-8.

[217] Statista: Life expectancy at birth in africa as of 2019, by country. https://www.statista.com/statistics/1218173/life-expectancy-in-african-countries/, 2021. Accessed: 14 November 2021.

[218] Coronavirus. http://coronavirus.politologue.com/coronavirus-Cameroun.cm, 2021. Accessed: 14 January 2021.

[219] H. Kawasuji, Y. Takegoshi, M. Kaneda, A. Ueno, Y. Miyajima, K. Kawago, Y. Fukui, Y. Yoshida, M. Kimura, H. Yamada, I. Sakamaki, H. Tani, Y. Morinaga, and Y. Yamamoto. Transmissibility of covid-19 depends on the viral load around onset in adult and symptomatic patients. *PLOS ONE*, 15(12), 2020. doi: 10.1371/journal.pone.0243597.

[220] S. E. Kim, H. S. Jeong, Y. Yu, S. Shin, S. Kim, T. H. Oh, U. J. Kim, S-J. Kang, H-C. Jang, S-I. Jung, and K-H. Park. Viral kinetics of sars-cov-2 in asymptomatic carriers and presymptomatic patients. *International Journal of Infectious Diseases*, 95: 441–443, 2020. doi: 10.1016/j.ijid.2020.04.083.

[221] B. R. Murphy, M. B. Rennels, R. G. Douglas, R. F. Betts, R. B. Couch, T. R. Cate, R. M. Chanock, A. P. Kendal, H. F. Maassab, S. Suwanagool, S. B. Sotman, L. A. Cisneros, W. C. Anthony, D. R. Nalin, and M. M. Levine. Evaluation of influenza a/hong kong/123/77 (h1n1)ts-1a2 and cold-adapted recombinant viruses in seronegative adult volunteers. *Infection and Immunity*, 29(2):348–355, 1980. doi: 10.1128/iai.29.2.348-355.1980.

[222] J. Demongeot, C. Jézéquel, and S. Sené. Boundary conditions and phase transitions in neural networks. theoretical results. *Neural Networks*, 21(7):971–979, 2008. doi: 10.1016/j.neunet.2008.04.003.

[223] J. Demongeot and S. Sené. Boundary conditions and phase transitions in neural networks. simulation results. *Neural Networks*, 21(7):962–970, 2008. doi: 10.1016/j.neunet.2008.04.005.

[224] J. Demongeot and S. Sené. Phase transitions in stochastic non-linear threshold boolean automata networks on z^2: The boundary impact. *Advances in Applied Mathematics*, 98:77–99, 2018. doi: 10.1016/j.aam.2018.03.003.

[225] R. Marchi, B. Sugita, A. Centa, A. S. Fonseca, S. Bortoletto, K. Fiorentin, S. Ferreira, and L. R. Cavalli. The role of micrornas in modulating sars-cov-2 infection in human cells: a systematic review. *Infection, Genetics and Evolution*, 91:104832–104832, 2021. doi: 10.1016/j.meegid.2021.104832.

[226] N. Golyandina and A. Korobeynikov. Basic singular spectrum analysis and forecasting with r. *Computational Statistics & Data Analysis*, 71:934–954, 2014. doi: 10.1016/j.csda.2013.04.009.

[227] Globaldatalab. https://globaldatalab.org/profiles/region/CMRr110/, 2021. Accessed: 14 November 2021.

[228] Organisation for economic co-operation and development. https://data.oecd.org/, 2021. Accessed: 14 November 2021.

[229] Africa sdg index and dashboards report. https://www.sdgindex.org/reports/2019-africa-sdg-index-and-dashboards-report/, 2021. Accessed: 14 November 2021.

[230] Adjusted net savings (% of gni) dimension: Socio-economic sustainability. http://hdr.undp.org/en/indicators/164406, 2021. Accessed: 14 November 2021.

[231] Environmental performance index. https://epi.yale.edu/, 2021. Accessed: 14 November 2021.

[232] J. Demongeot, N. Noury, and N. Vuillerme. Data fusion for analysis of persistence in pervasive actimetry of elderly people at home, and the notion of biological age. *2008 International Conference on Complex, Intelligent and Software Intensive Systems*, 2008. doi: 10.1109/cisis.2008.30.

[233] S. Soubeyrand, J. Demongeot, and L. Roques. Towards unified and real-time analyses of outbreaks at country-level during pandemics. *One Health*, 11:100187–100187, 2020. doi: 10.1016/j.onehlt.2020.100187.

[234] F. Ferraty and P. Vieu. Curves discrimination: a nonparametric functional approach. *Computational Statistics & Data Analysis*, 44(1):161–173, 2003. doi: 10.1016/s0167-9473(03)00032-x.

[235] F. Ferraty and P. Vieu. Nonparametric models for functional data, with application in regression, time series prediction and curve discrimination. *Journal of Nonparametric Statistics*, 16(1):111–125, 2004. doi: 10.1080/10485250310001622686.

[236] F. Ferraty and P. Vieu. Functional nonparametric statistics: a double infinite dimensional framework. *Recent Advances and Trends in Nonparametric Statistics*, pages 61–76, 2003. doi: 10.1016/b978-044451378-6/50005-3.

[237] J. Demongeot, A. Laksaci, F. Madani, and M. Rachdi. Functional data: local linear estimation of the conditional density and its application. *Statistics*, 47(1):26–44, 2013. doi: 10.1080/02331888.2011.568117.

[238] M. Rachdi, A. Laksaci, J. Demongeot, A. Abdali, and F. Madani. Theoretical and practical aspects of the quadratic error in the local linear estimation of the conditional density for functional data. *Computational Statistics & Data Analysis*, 73:53–68, 2014. doi: 10.1016/j.csda.2013.11.011.

[239] J. Demongeot, A. Laksaci, M. Rachdi, and S. Rahmani. On the local linear modelization of the conditional distribution for functional data. *Sankhya A*, 76(2):328–355, 2014. doi: 10.1007/s13171-013-0050-z.

[240] J. Demongeot, A. Hamie, A. Laksaci, and M. Rachdi. Relative-error prediction in nonparametric functional statistics: Theory and practice. *Journal of Multivariate Analysis*, 146:261–268, 2016. doi: 10.1016/j.jmva.2015.09.019.

[241] J. Demongeot, A. Naceri, A. Laksaci, and M. Rachdi. Local linear regression modelization when all variables are curves. *Statistics & Probability Letters*, 121:37–44, 2017. doi: 10.1016/j.spl.2016.09.021.

[242] B. Altendji, J. Demongeot, A. Laksaci, and M. Rachdi. Functional data analysis: estimation of the relative error in functional regression under random left-truncation model. *Journal of Nonparametric Statistics*, 30(2):472–490, 2018. doi: 10.1080/10485252.2018.1438609.

[243] A. Henien, L. Ait-Hennani, J. Demongeot, A. Laksaci, and M. Rachdi. Test d'hétéroscédasticité quand les covariables sont fonctionnelles. *Comptes Rendus Mathematique*, 356(5):571–574, 2018. doi: 10.1016/j.crma.2018.02.010.

[244] Données hospitalières relatives à l'épidémie de covid-19, 2021. https://www.data.gouv.fr/fr/datasets/donnees-hospitalieres-relatives-a-lepidemie-de-Covid-19, 2021. Accessed: 14 November 2021.

[245] Live covid-19 vaccination tracker. https://covidvax.live/location/fra, 2021. Accessed: 14 November 2021.

[246] A. Srivastava and E. P. Klassen. Functional data and elastic registration. *Functional and Shape Data Analysis*, pages 73–123, 2016. doi: 10.1007/978-1-4939-4020-2_4.

[247] B. Tang, X. Wang, Q. Li, N. L. Bragazzi, S. Tang, Y. Xiao, and J. Wu. Estimation of the transmission risk of the 2019-ncov and its implication for public health interventions. *Journal of Clinical Medicine*, 9(2):462–462, 2020. doi: 10.3390/jcm9020462.