STATISTICAL MODEL IN ESTIMATION AND MONITORING OF COVID-19

Thesis submitted to

THE KLE ACADEMY OF HIGHER EDUCATION AND RESEARCH, BELAGAVI
(KLE DEEMED UNIVERSITY)

[Declared as Deemed-to-be-University u/s 3 of the UGC Act, 1956 vide Govt. of India Notification No.F.9-19/2000-U.3 (A)]
(Accredited 'A+' Grade by NAAC) (3rd Cycle)
[Placed in Category 'A' by MoE(GoI)]

For the award of the degree of

Doctor of Philosophy in the Faculty of Interdisciplinary Science (Biostatistics)

By

MR. NOEL GEORGE

(Registration No: KAHER/Ph.D./20-21/DO1220059)

Under the Guidance of

DR. J. B. PRASAD

Associate Professor & I/C HOD,
DEPARTMENT OF EPIDEMIOLOGY & BIOSTATISTICS,
KAHER, BELAGAVI

December- 2023

UNDERTAKING

I, **Mr. Noel George** hereby declare that the information and the data mentioned in my thesis entitled **"Statistical model in estimation and monitoring of COVID-19"** belongs to me and is original.

I am aware of definition of plagiarism as detailed below:

- An act or instance of using or closely imitating the language and thoughts of another author without authorization and the representation of that author's work as one's own, as by not crediting the original author.
- A piece of writing or other work reflecting such unauthorized use or imitation.
- The deliberate or reckless representation of another's words, thoughts or ideas as one's own without attribution in connection with submission of academic work, whether graded or otherwise.

I hereby declare that the thesis prepared by me is original-one and does not involve plagiarism anywhere. In case at a later stage it is found that I have indulged in plagiarism, then I am solely responsible for the same and the Institution is at liberty to take any disciplinary action against me including cancellation of dissertation or any other penalties imposed by the University.

Noel George
Research Scholar

Date: 14/12/23

Place: Belagavi

PLAGIARISM CERTIFICATE

KLE ACADEMY OF HIGHER EDUCATION AND RESEARCH
(Formerly known as KLE University)
(Deemed-to-be-University established u/s 3 of the UGC Act, 1956)
Accredited A+ Grade by NAAC (3rd Cycle) Placed in Category 'A' by MoE (GoI)
JNMC Campus, Nehru Nagar, Belagavi-590 010, Karnataka State, India
☎: 0831-2444444 Web: http://www.kledeemeduniversity.edu.in E-mail: info@kledeemeduniversity.edu.in

Ref. No. KAHER/AA/23-24/D- 314 16th November 2023

Sir,

The soft copy of Ph.D. research thesis of **Mr. Noel George, Faculty of Interdisciplinary Science, KAHER, Belagavi** has been submitted for anti-plagiarism check at the office of the undersigned through "Turn-it-in" package. The scan has been carried out and the scanned output reveals a match percentage of **5%** which is within the acceptable limit of 10%.

To obtain the comprehensive report of the plagiarism test, research scholar can send a mail to **diracademic@kledeemeduniversity.edu.in** along with the Registration Number, Name of the Scholar, Name of Guide/Co-guide and title of the thesis.

DIRECTOR ACADEMIC AFFAIRS Deemed-to-be University

Dr.(Mrs.) Roopa M. Bellad
Director, Academic Affairs

To,

Mr. Noel George
Full-Time Ph.D. Scholar,
2020-21 Batch,
Faculty of Interdisciplinary Science,
KAHER, **Belagavi.**

Cc to :
1. The Principal, JNMC, KAHER, Belagavi
2. Dr. Jang Bahadur Prasad, Associate Prof., of Epidemiology & Biostatistics, KAHER, Belagavi - Guide

KLE ACADEMY OF HIGHER EDUCATION AND RESEARCH
(Deemed-to-be-University)
[Declared as Deemed-to-be-University u/s 3 of the UGC Act, 1956 vide Govt. of India Notification No.F.9-19/2000-U.3 (A)]
Accredited 'A^{+}' Grade by NAAC (3rd Cycle)
Placed in Category 'A' by MoE (GoI)

KLE
EMPOWERING PROFESSIONALS

COPYRIGHT DECLARATION

We hereby declare that **KLE ACADEMY OF HIGHER EDUCATION AND RESEARCH, BELAGAVI, KARNATAKA,** *shall have the rights to preserve, use and disseminate this thesis in print or electronic format for academic/research purpose.*

Noel George
Research Scholar

Dr. J. B. Prasad,
Associate Professor & I/C HOD
Department of Epidemiology & Biostatistics, KAHER

Place: Belagavi

Date: 14/12/23

DEPARTMENT OF EPIDEMIOLOGY & BIOSTATISTICS • KAHER'S
BELAGAVI

KLE ACADEMY OF HIGHER EDUCATION AND RESEARCH
(Deemed-to-be-University)
[Declared as Deemed-to-be-University u/s 3 of the UGC Act, 1956 vide Govt. of India Notification No.F.9-19/2000-U.3 (A)]
Accredited 'A^{+}' Grade by NAAC (3rd Cycle)
Placed in Category 'A' by MoE (GoI)

KLE
EMPOWERING PROFESSIONALS

DECLARATION

*I hereby declare that the thesis entitled "***Statistical model in estimation and monitoring of COVID-19"** is a bonafide and original research carried out by me under the guidance of **Dr. J. B. Prasad,** Associate Professor & I/C HOD, Dept. of Epidemiology & Biostatistics, KAHER.

The thesis or any part thereof has not formed the basis for the award of any degree/fellowship or similar title to any candidate of any University.

Place: Belagavi
Date: 14/12/23

Noel George
Ph.D. Scholar
Department of Epidemiology
& Biostatistics, KAHER

KLE ACADEMY OF HIGHER EDUCATION AND RESEARCH
(Deemed-to-be-University)
[Declared as Deemed-to-be-University u/s 3 of the UGC Act, 1956 vide Govt. of India Notification No.F.9-19/2000-U.3 (A)]
Accredited 'A^{+}' Grade by NAAC (3rd Cycle)
Placed in Category 'A' by MoE (GoI)

KLE
EMPOWERING PROFESSIONALS

CERTIFICATE

This is to certify that the thesis entitled "**Statistical model in estimation and monitoring of COVID-19"** is a bonafide and genuine research carried out by **Mr. Noel George** under the guidance of **Dr. J. B. Prasad,** Associate Professor & I/C HOD, Dept. of Epidemiology & Biostatistics, KAHER.

Place: Belagavi
Date: 14/12/23

Dr. R. B. Nerli
Dean
Interdisciplinary Sciences
JNMC, KAHER

Dr. R. B. NERLI
MS.,Mch.,Ph.D.,MBA
Dean
Interdisciplinary Sciences
KAHER

KLE ACADEMY OF HIGHER EDUCATION AND RESEARCH
(Deemed-to-be-University)
[Declared as Deemed-to-be-University u/s 3 of the UGC Act, 1956 vide Govt. of India Notification No.F.9-19/2000-U.3 (A)]
Accredited 'A^{+}' Grade by NAAC (3rd Cycle)
Placed in Category 'A' by MoE (GoI)

KLE
EMPOWERING PROFESSIONALS

CERTIFICATE

This is to certify that the thesis entitled "**Statistical model in estimation and monitoring of COVID-19**"is a bonafide record of original research carried out by **Mr. Noel George** for the award of degree of DOCTOR OF PHILOSOPHY IN FACULTY OF INTERDISCIPLINARY SCIENCE (BIOSTATISTICS) under my supervision and guidance

Place: Belagavi
Date: 14/12/23

Dr. J. B. Prasad
Associate Professor & I/C HOD,
Dept. of Epidemiology & Biostatistics,
KAHER, Belagavi

DEPARTMENT OF EPIDEMIOLOGY & BIOSTATISTICS • KAHER'S
BELAGAVI

ACKNOWLEDGMENT

Life provides limited opportunities to express heartfelt thanks and appreciation to those indispensable in achieving our goals. I eagerly seize this chance to extend my profound gratitude for the assistance, support, encouragement, and guidance, I received during this thesis's completion. Above all, I am deeply grateful to my parents **(Mr. George Joseph and Mrs. Honey George)** for their countless sacrifices throughout my educational journey. To my parents, whose unwavering support and love have been my constant source of strength, I owe an immense debt of gratitude.

I am deeply grateful for the guidance and support provided by **Dr. Jang Bahadur Prasad** from the Department of Epidemiology & Biostatistics at KLE Academy of Higher Education and Research. His expertise and insights have been invaluable to my research.

I would like to express my heartfelt gratitude to the esteemed **Vice Chancellor Prof (Dr.) Nitin M. Gangane,** and the Ph.D. expert committee at KLE Academy of Higher Education and Research. Their insightful feedback was invaluable in the successful completion of my research.

Additionally, my acknowledgments are to **Registrar Prof. (Dr.) M. S. Ganachari** at KAHER, and **Former Registrar Prof. (Dr.) V. A. Kothiwale** for their support in providing the necessary institutional facilities throughout my research journey.

I am obliged to **Dr. Roopa Bellad, Director of Academic Affairs,** KAHER, **Mrs. Swati Samuel,** Officer in charge International Admissions, KAHER,

Academic Affairs, Department and for providing all the necessary resources in completion of my thesis and supporting during course.

I would also like to express my appreciation to the former Head of the Department, **Dr. Naresh Kumar Tyagi,** for his foundational work and contributions. Special thanks are extended to **Dr. Rajeshwari Biradar, Ms. Anushri P. Patil, Dr. Sudhirgouda Patil** and **Mrs. Roopa Chougule** for their assistance and dedication throughout this research journey.

I am honoured to acknowledge Emeritus **Professor K. K. Jose, M.Sc.,** M.Phil., Ph.D., FRSS, FSMS, Hon. Director of the School of Mathematics & Statistics at MGU, for his inspirational guidance and academic wisdom. My gratitude goes to the members and supporters from the Department of Biostatistics at St. Thomas College, Palai, whose encouragement and support have been pivotal.

I am thankful to my friends and colleagues for their camaraderie and constant support throughout this endeavour.

Finally, I extend my thanks to all who have directly and indirectly contributed to the completion of my thesis work. Your support has been a cornerstone of my achievements. Their collective wisdom, guidance, and support have been the pillars of strength in my academic journey.

Date: **Signature of the Research Scholar**

Place: Belagavi **Noel George**

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ACF	-	Autocorrelation Function
ADF	-	Augmented Dickey-Fuller test
AIC	-	Akaike Information Criterion
AQI	-	Air Quality Index
AR	-	Auto Regressive model
ARIMA	-	Auto Regressive Integrated Moving Average
ARMA	-	Auto Regressive Moving Average
BIC	-	Bayesian Information Criterion
CDF	-	Cumulative Distribution Function
CFR	-	Case Fatality Rate
CI	-	Confidence Interval
CMR	-	Crude Mortality Rate
CO	-	Carbon Monoxide
COVID	-	Coronavirus Disease
CPCB	-	Central Pollution Control Board
CSS	-	Conditional Sum of Squares
CSSE	-	Center for Systems Science and Engineering
DASS	-	Depression Anxiety Stress Scales
DR	-	Death Rate
DVR	-	Data Validation and Reconciliation
ICU	-	Intensive Care Unit
IID	-	Independent and Identically Distributed
ITS	-	Interrupted Time Series

JHU	-	Johns Hopkins University
KPSS	-	Kwiatkowski-Phillips-Schmidt-Shin (test)
LM	-	Lagrange Multiplier
LOF	-	Lack of Fit
MA	-	Moving Average
MAE	-	Mean Absolute Error
MAPE	-	Mean Absolute Percentage Error
MERS	-	Middle East Respiratory Syndrome
MLE	-	Maximum Likelihood Estimation
MR	-	Mortality Rate
ORAQI	-	Operational Real-time Air Quality Index
ORNAQI	-	Operational Real-time National Air Quality Index
ORNL	-	Oak Ridge National Laboratory
PACF	-	Partial Autocorrelation Function
PCR	-	Polymerase Chain Reaction
PDF	-	Probability Density Function
PM	-	Particulate Matter
PPE	-	Personal Protective Equipment
QQ	-	Quantile-Quantile
RNA	-	Ribonucleic Acid
RR	-	Relative Risk
RSS	-	Residual Sum of Squares
SARS	-	Severe Acute Respiratory Syndrome
SE	-	Standard Error
SEIR	-	Susceptible Exposed Infectious Recovered (model)

SIR	-	Susceptible Infectious Recovered (model)
SIRD	-	Susceptible Infectious Recovered Deceased (model)
SWOT	-	Strengths, Weaknesses, Opportunities, Threats
TSS	-	Total Sum of Squares
UHC	-	Universal Health Coverage
USA	-	United States of America
WHO	-	World Health Organization

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Background: In the last two decades, the world has seen several viral outbreaks, notably the H1N1 in 2009 and Nipah virus in 2018. The COVID-19 pandemic, declared by WHO in early 2020, presented unprecedented challenges due to its high contagion and potential for asymptomatic spread. Early data often underestimated the pandemic's extent, emphasizing the need for comprehensive health crisis management. Statistical models have been crucial in understanding and responding to COVID-19, from tracing its origins in Wuhan, China, to its global spread. These models have informed public health strategies, adapting to regional differences in the pandemic's impact.

The significance of this study lies in using statistical models to estimate and monitor COVID-19, aiding policymakers and healthcare professionals. It sheds light on the virus's transmission and future trends, guiding targeted actions and preparedness for future pandemics. The research underscores the importance of current data over historical records, considering behavioural changes, new virus strains, and varying testing/reporting methods.

Objectives: The objectives of this study are to establish a model for the COVID-19 pandemic and provide effective methods, enabling planners and healthcare providers to take appropriate actions for the management, containment, and control of the pandemic.

Materials and Methods: A comprehensive and data-driven approach is employed to unravel the intricate dynamics of the COVID-19 pandemic. The methodology incorporates a variety of tools and techniques designed to capture the multifaceted

aspects of the virus's spread and impact. The initial phase involves an extensive data collection process, drawing from diverse sources, including open-source COVID-19 databases and supplementary data repositories, ensuring the acquisition of a broad and inclusive dataset that forms the foundation for subsequent analyses. Following data collection, a meticulous data curation process is undertaken, incorporating pre-processing, reconciliation, and fusion techniques to enhance the quality and reliability of the gathered data, ensuring the dataset is robust and reflective of the real-world scenario.

The analytical methods include fundamental techniques such as clustering, real-time surveillance, and time-series theory, providing a solid framework for understanding the patterns and trends in the progression of the pandemic, offering insights that inform subsequent decision-making. To monitor and evaluate the evolving situation, techniques like SWOT analysis, consolidated time-series analysis, and model selection criteria are employed. These monitoring techniques contribute to a holistic assessment of the pandemic dynamics, allowing for a nuanced understanding of its various facets. The modelling approaches adopted in this study include Susceptible-Infectious-Recovered (SIR) and exponential growth models, serving as essential tools for estimation and prediction. This integration enhances the study's predictive capabilities and supports evidence-based decision-making.

Result: The research provides a multi-faceted analysis of the COVID-19 pandemic focusing in India, employing time-series analysis and growth models such as Exponential, Logistic, Gompertz, and Bertalanffy to predict the virus's overall spread. The basic reproduction number (R_0) in the entire country is estimated to be 1.01 and 1.11 days for SIR and SIRD models, respectively, within the initial 120 days of the

pandemic, indicating that these models are best suited for prediction in the early phase of the outbreak.

The effectiveness of lockdown measures is evaluated through trends across different Indian states, with time-series analysis and hierarchical clustering tools. The average recovery time from COVID-19 varied across Indian states, with most states falling between 5 to 36 days, except Madhya Pradesh. The study also notes a lack of significant correlation in the virus spread patterns between different countries and discusses spatial epidemiology and forecasting.

This comprehensive assessment indicates the interplay between health, environmental, and socioeconomic factors in managing the pandemic. The lockdown led to a 37% reduction in air pollutants, showcasing a significant environmental impact. The analysis of COVID-19 data for states like Tamil Nadu and Maharashtra reveals distinct patterns in the active cases and forecasting models. In Tamil Nadu, a model of (0,2,1) is identified based on autocorrelation and partial autocorrelation functions, with residuals demonstrating no significant correlation and passing normality tests. In-sample forecasting using this model indicates reasonably accurate predictions for a few upcoming days. Similarly, Maharashtra's analysis employs an ARIMA (2,2,0) model, showing uncorrelated residuals and passing normality tests. Forecasting active cases in Maharashtra using this model suggests a high degree of accuracy for the specified period.

The study further explores the demand and capacity of hospital infrastructure in these states, indicating potential challenges in managing ICU, ventilator, and bed capacities. Additionally, a social impact assessment reveals a high level of awareness about precautionary measures, with a notable psychological impact on the population.

The association between demographic variables and psychological impact, anxiety, stress, and depression is explored, providing insights into the diverse social and psychological effects of the pandemic. The lockdown also resulted in increased stress levels among individuals with lower monthly family incomes, and the sources from which people obtained information about the COVID-19 outbreak significantly influenced the rise in stress levels during the pandemic.

Conclusion: Different statistical models were fitted, and the models for indenting the average recovery rates of the COVID-19 patients found that Tamil Nadu was the best fit with an R^2 value of 0.96, and Mizoram was identified to have the least fit with an R^2 value of 0.58. At the same time, Manipur was identified to have the least average recovery duration with 4.94 days and the highest in Madhya Pradesh with 68 days. The first and second COVID-19 wave in India had different characteristics. The second phase had almost five times more when compared with the first. As of May, 2021, the resurgent wave of COVID-19 in India has been spreading to rural populations. Yet in the present moment, India confronts an urgent need to save lives and alleviate suffering.

Keywords: COVID-19 pandemic, Statistical models, Estimation, Monitoring, Making Decision

CHAPTER 1: INTRODUCTION

1.1 Background

1.1.1. COVID-19: Origin and Spread

1.1.1.1. Origin of COVID-19

The World Health Organisation (WHO) on December 31, 2019, received a press release from the official website of the Wuhan Municipal Health Commission, confirming the existence of a group of flu cases in Wuhan, People's Republic of China.[1] These cases exhibited symptoms similar to those of 'viral pneumonia'.[2] However, at that time, the global community was unaware that these initial cases were connected to the COVID-19 infection, as seen in the accompanying image.[3] This virus would soon cause widespread devastation, significantly impacting and disrupting the lives of people worldwide.[4] The emergence of COVID-19 led nations to implement various forms of lockdowns and restrictions, resulting in a substantial alteration of daily routines, economies, and societal norms. The consequences of this pandemic have been substantial and long-lasting, fundamentally changing how people live, work, and interact with one another.[5]

SARS-CoV-2, the virus responsible for causing COVID-19, was named after the virus, SARS-CoV-1, which caused a severe outbreak in 2002-2003, known as severe acute respiratory syndrome (SARS).[6] Viruses are intriguing entities; they are not classified as living organisms, yet they possess genetic material that enables them to take control of the cellular machinery of host cells and reproduce themselves. This ability allows viruses to generate multiple copies of their genetic material, facilitating the spread of infection.[7] The term "host-switching" describes a phenomenon in which a disease is transmitted between different species, often involving a vertebrate and an

insect.[8] Many infectious diseases, both viral and non-viral, that have affected humans over centuries, originated from animal-to-human transmissions. The study of animal viruses found in humans provides valuable insights into comprehending the factors and mechanisms that drive host-switching events.[9]

The phenomenon of viruses transitioning from animals to humans, known as zoonotic transmission, is widely recognized with the influenza virus serving as a prominent illustration. Influenza viruses affect humans, encompassing both pandemic and seasonal strains, which are traced back to enzootic viruses naturally found in wild waterfowl and shorebirds.[10] These avian viruses typically coexist benignly with their avian hosts, causing no severe illnesses. Nevertheless, under specific conditions, they undergo genetic alterations that enable them to traverse species boundaries and infect mammals, including humans.[11] Coronaviruses, another category of RNA viruses, usually induce mild upper respiratory tract infections in humans. The historical origins of these viruses continue to be a subject of ongoing investigation. Nevertheless, it is postulated that coronaviruses have coevolved alongside humans and may have conferred certain benefits to our immune systems.[12]

The virus responsible for the COVID-19 global outbreak, SARS-CoV-2, has been identified as the third bat-borne virus associated with the emergence of a human disease, demonstrating the significance of zoonotic events in public health. This virus is believed to have its origin in bats, where it circulates naturally without causing severe illness.[13] However, through a process of zoonotic transmission, it was introduced to humans, precipitating a global pandemic that has had profound consequences on societies and individuals worldwide.[14] The exploration of zoonotic viruses, exemplified by SARS-CoV-2, holds paramount importance as it allows for

the elucidation of the mechanisms involved in the spillover from animal reservoirs to the human population.[15] By acquiring a deeper understanding of the factors facilitating cross-species transmission, the scientific community enhances preparedness and response measures for future outbreaks. This knowledge empowers researchers to develop strategies aimed at averting and mitigating the impact of emerging infectious diseases on a global scale.[16]

In the year 2016, a novel epizootic ailment affecting swine populations in China was determined to be the result of a coronavirus. Subsequently, in late November 2019, the emergence of SARS-CoV-2, the virus responsible for the COVID-19 pandemic, was confirmed.[17] Bats, more prominently than other mammals, are recognised as a natural reservoir for coronaviruses.[18] A comprehensive research endeavour encompassing the investigation of over 19,000 animals across 20 countries revealed that an overwhelming 98% of coronavirus cases were associated with bats. Both SARS-CoV-2 and its predecessor, SARS-CoV-1, trace their origins to rhinophid viruses existing in the natural environment.[19] Researchers have undertaken mapping initiatives aimed at identifying global hotspots where novel infections could potentially manifest, with a specific emphasis on regions situated in southern and southwestern China.[20]

Upon scrutinising the genetic makeup of coronaviruses, scientists made a noteworthy discovery: the genome of one of these viruses exhibits an astonishing 96% similarity to SARS-CoV-2. Notably, distinctive features such as the furin cleavage site insertion and a genetic resemblance exceeding 97% in the lab replicase gene have been identified between these two viruses. These results offer important perspectives on the evolutionary lineage and potential origins of SARS-CoV-2, underscoring the

critical importance of comprehending the dynamics of zoonotic transmission in the context of managing and pre-empting future infectious disease outbreaks.[21] The ongoing pursuit of research in this domain holds the promise of fortifying preparedness measures and bolstering public health responses when confronting emerging viral threats.

A "hotspot" for bat coronaviruses refers to a specific region in south and southwest China, which has been identified as an area with a high prevalence of these viruses.[22] Coronaviruses have been known to utilise human receptors for angiotensin-converting enzyme 2 (ACE2) as targets, making them capable of causing pandemics in humans.[23] Initial investigations have revealed the potential of certain coronaviruses to spread widely among human populations.[24] Distinct genetic fingerprints have been observed in SARS-CoV-2, the virus responsible for COVID-19, which set it apart from previous coronavirus engineering attempts or "forward engineered" sequences derived from closely related sarbecoviruses.[25] The unique genetic makeup of SARS-CoV-2 indicates that it is not a product of known coronavirus manipulation efforts. These results underscore the necessity for additional studies to comprehend the genesis and spread patterns of SARS-CoV-2.[26] The absence of direct evidence for laboratory engineering or manipulation indicates that the virus likely emerged naturally through zoonotic transmission. As investigations continue, it is crucial to gain deeper insights into the evolutionary history of SARS-CoV-2 to better prepare for and prevent future outbreaks of novel coronaviruses.[27]

1.1.1.2. Global Spread of COVID-19

A worldwide outbreak caused by the COVID-19 virus has emerged, recording more than 25 million verified infections and surpassing 250,000 deaths globally.[28] The virus initially surfaced in Asia but subsequently migrated, impacting regions in South America, North America, and Europe, with a recent resurgence in North America.[29] Nations worldwide swiftly mobilised to establish diverse diagnostic methodologies for identifying individuals displaying COVID-19 symptoms. As testing became more widely accessible, the tally of infected and deceased individuals attributed to the virus was meticulously documented across various systems, facilitating the construction of statistical models.[30]

China officially recorded its first coronavirus-related fatality on January 11, marking over a million lives lost to the disease eight months after its initial identification.[31] Specifically in Wuhan, the virus responsible for COVID-19, SARS-CoV-2, spread swiftly, resulting in a thousand deaths within just one month.[32] This initial mortality rate exceeded the 774 fatalities attributed to the earlier respiratory disease, SARS, during its 2002-2003 outbreak in Asia.[33] While most countries and territories beyond mainland China remained relatively unaffected at that time, the virus had already disseminated globally.[34] The Philippines reported its first COVID-19 case on February 2, 2020, followed by Hong Kong two days later, Japan on February 13, 2020, and France on February 14, 2020.[35] The month of February witnessed a substantial surge in reported cases. By March 11, 2020, the World Health Organisation (WHO) declared the novel coronavirus a "pandemic," following the identification of 4,500 deaths in 30 countries and territories. The majority of these

deaths remained concentrated in China, though Iran (with 300 deaths) and Italy (with 800 deaths) witnessed rapid escalations in incidents.[36]

The daily death toll in the United States and Europe continued to rise until mid-April, peaking at an average of over 2,700 and 4,000 deaths per day, respectively, during the second week of April.[37] The United States presently maintains its status as the worst-affected nation globally, with approximately 200,000 deaths.[38] The direst week on a global scale unfolded during the period from April 13 to 19, 2020, recording an average of around 7,460 daily COVID-19 related deaths. By that time, the global death toll had climbed to nearly 170,000, more than double the figure reported on March 31, 2020. Since the outset of June, there has been no discernible alteration in the daily average of fatalities.[39] Latin America and the Caribbean assumed the mantle of the new epicentre of the pandemic in June, with daily recorded deaths in the region consistently exceeding 2,500 between July 15 and August 15, 2020.[40]

Subsequently, a modest downturn in numbers materialised, with an average of 1,900 daily deaths reported by September 2020.[41] Brazil overtook the United States as the country with the highest overall death toll, exceeding 140,000. Countries such as Peru (975) and Bolivia (671), as well as European nations like Spain (671) and Belgium (861), have been among the hardest-hit nations per capita.[42] In Asia, the daily death toll averaged less than one hundred individuals until mid-April but has steadily risen since then. Since July 20, 2020, the continent has reported an average of over 1,000 daily deaths, approaching 1,500. India, in particular, has borne a significant brunt, with over 480,000 lives lost due to the virus.[43]

By the end of 2020, new cases had risen by approximately 20% across the continent, while fatalities had surged by 28%.[44] Analogous to other regions globally, the Middle East has witnessed a resurgence in daily fatalities, with an increase of around 18% over two weeks, tallying around 330 deaths.[45] In contrast, Africa has been relatively less impacted, with fatalities declining since August to less than 200 daily deaths by mid-September, following an earlier peak of around 400 daily deaths in August. Oceania, on the other hand, has consistently maintained daily death tolls below twenty-four.[46]

The first patient was admitted on December 12, 2019, followed by the initial Chinese national succumbing to the virus at the beginning of January 2020, cases linked to COVID-19 through travel began emerging in various countries, including Japan, Thailand, France, South Korea, and the United States of America (USA) by mid-January 2020.[47] By the end of January 2020, the virus had already disseminated across the West Pacific, Southeast Asia, Europe, Canada, the United States, and the East Mediterranean, precipitating a critical global public health emergency.[48] As of June 9, 2020, there were 7,039,918 confirmed cases, 404,396 fatalities, and 3,596,972 recoveries worldwide. Projections indicate that the cumulative count of COVID-19 cases may reach approximately 272 million by December 15, 2021, with the United States accounting for roughly one-fifth of all reported cases globally, demonstrating the virus's pervasive reach.[49]

The COVID-19 outbreak has exacted a substantial human toll worldwide, posing unprecedented challenges to food systems, public health infrastructure, and workplaces.[50] The economic and social reverberations of the pandemic were profound, pushing tens of millions into poverty, with the number of undernourished

individuals projected to have risen to 132 million by the end of the year 2020.[51] The pandemic's implications have exacerbated the risk of extreme poverty for many. It is clear that the COVID-19 pandemic has cast a wide net of repercussions, affecting various facets of human life and underscoring the imperative for robust public health measures and global cooperation to confront and mitigate future outbreaks effectively.[52]

1.1.1.3. Government Measures

The effective implementation of vital government measures, such as public health campaigns, healthcare infrastructure improvement, and the implementation of lockdowns has been instrumental during the pandemic. These actions highlight the government's commitment to safeguarding public health and emphasise the importance of a resilient, adaptable, and responsive public health system.[53] Furthermore, they underscore the critical role of rational procurement policies, procedures, and systems in ensuring the swift acquisition of life-saving resources during emergencies like the COVID-19 crisis. These comprehensive measures encompassed a range of strategies, including mobility surveillance, physical distancing, mobility restrictions, quarantine protocols, proactive testing, and contact tracing, among others.[54] These strategies played a pivotal role in effectively managing and responding to the numerous challenges posed by the pandemic.

When the COVID-19 pandemic first emerged, there was a significant lack of established guidelines, given the novelty of the virus.[55] The World Health Organization (WHO) responded by drawing on previous epidemic control strategies and adapting them to the unique circumstances of the crisis. Their efforts highlighted the importance of learning from past experiences to inform current responses.[56]

Governments around the world grappled with the unprecedented challenge of responding to the pandemic, and each nation took its own approach based on its specific circumstances and resources. This decentralised response showcased the need for flexibility and adaptability in the face of a global crisis.[57] The impact of the pandemic varied widely from country to country, depending on their approach and circumstances. For instance, nations like New Zealand, with relatively closed borders, implemented strict containment measures that proved effective in curbing the virus's spread. In contrast, countries like India, with more open borders and densely populated cities, faced unique challenges in controlling the outbreak.[58]

The COVID-19 pandemic underscored the absence of a one-size-fits-all solution to such a global crisis. Instead, it highlighted the need for diverse models and strategies, informed by statistical data and tailored to individual circumstances, to effectively control the pandemic.[59] Governments and healthcare systems around the world faced unique challenges and made crucial decisions based on the available evidence and resources. As we transitioned into the next phase, the focus on vaccination emerged as a pivotal strategy in our collective efforts to combat the virus and ultimately restore a semblance of normalcy.[60]

1.1.1.4. Vaccination

Vaccination stands as a critical pillar in the global battle against COVID-19, playing a pivotal role in curbing the virus's transmission, preventing severe illness, and saving lives.[61] These campaigns have significantly bolstered immunity levels within populations, facilitating a gradual return to normalcy and the pandemic's containment. Ongoing research and efforts persist in refining vaccination strategies, addressing virus variants, and ensuring equitable global vaccine access, underscoring

their paramount role in our collective response to this unparalleled global health challenge.[62] The advent of COVID-19 vaccines has instilled a sense of hope in our pandemic response. These vaccines mark a significant milestone, offering the promise of widespread immunity and a gradual return to pre-pandemic life.[63]

In the comprehensive examination of strategies to combat the COVID-19 pandemic, it is crucial to emphasise the sequential nature of pandemic response planning.[64] Recognising vaccination as a secondary measure, the initial phase of the study is dedicated to the establishment of resilient control mechanisms rooted in statistical models and leveraging available data.[65] Consequently, the present investigation intentionally excludes vaccination as a variable to concentrate on a meticulous assessment of the inherent spread, impact, and potential mitigation strategies for COVID-19. By delineating the phased approach to pandemic control, the study underscores the primary focus on developing models based on existing statistics to comprehend and address the complex dynamics of the ongoing pandemic. The deliberate omission of vaccination as a variable underscores the significance of statistical models in shaping understanding and formulating effective responses to the evolving challenges posed by COVID-19.

1.1.1.5. Pandemic Management

In the historical context of epidemiology, an illustrative example highlights the significant role of statistics in managing pandemics.[66] This was the first time a statistical model was being used for pandemic management. John Snow's work in mapping and analysing the spread of cholera in 19th-century London marks a significant shift in how we deal with infectious diseases during pandemics.[67] The use of statistical methods in this context has set the stage for future developments in

epidemiology, showing the potential of data-driven approaches in guiding interventions and understanding how diseases spread.[68]

In contemporary pandemic management, statistical models serve as a crucial tool in navigating the complexities of healthcare, policy-making, and public awareness.[69] The decision-making process, from formulating social distancing guidelines to strategically distributing vaccines, relies on the empirical foundation provided by statistical analyses. The historical legacy of epidemiological statistics seamlessly extends into the modern era, leveraging technological advancements to enhance statistical modelling.[70]

Epidemiological modelling, situated at the forefront of pandemic response, exemplifies the utility of statistical tools.[71] Models such as SIR (Susceptible, Infected, Recovered) and SEIR (Susceptible, Exposed, Infected, Recovered) assist in understanding disease spread dynamics, providing forecasts for infections, recoveries, and fatalities.[72] Modern statistical methods, incorporating technology-driven approaches, explore techniques for outbreak prediction and identifying factors influencing transmission dynamics.[73] In the intricate process of allocating pandemic resources, statistical optimisation algorithms emerge as indispensable. Faced with the scarcity of resources like medical personnel, ventilators, and vaccines, these algorithms guide the judicious distribution of critical assets. Analytics, informed by statistical insights, identify regions with escalating cases, enabling swift resource mobilisation to areas in greatest need.[74]

Furthermore, statistics play a role in prioritising vaccine distribution, ensuring vulnerable populations are addressed based on empirical risk assessments.[75] Beyond resource deployment, effective communication of pandemic-related information relies

on accurate statistical data. Visualisation tools such as charts, graphs, and heat maps simplify complex information, promoting a deeper understanding among the general population.[76] Statistics also influence public sentiment, mitigating unnecessary panic or instilling a necessary sense of urgency through an informed understanding of data significance. In the evolving landscape of infectious diseases, statistics serve as the empirical foundation for evidence-based decision-making in pandemic management. From modelling virus spread to resource allocation and public communication, statistical methods continue to be indispensable in navigating the complex challenges posed by public health crises.[77]

1.1.1.6. Challenges Faced in Estimation and Monitoring

Accurate estimation and effective monitoring of such events are crucial for timely intervention and containment.[78] Despite advancements in technology and healthcare systems, the task of estimating and monitoring pandemics faces multiple obstacles.[79] Understanding these challenges is pivotal for improving response mechanisms and public health outcomes. One of the foremost challenges in pandemic monitoring is data inconsistency and availability. The reliability of data varies widely between countries and even within regions, due to differences in testing capabilities, reporting systems, and healthcare infrastructure. Inaccurate or incomplete data not only skew estimations but also affect the allocation of resources and public health interventions.[80]

Public compliance with health directives is essential for effective monitoring. However, misinformation, socio-political factors, and distrust in governmental agencies undermine public compliance, thereby compromising the quality of epidemiological data and the effectiveness of containment measures.[81] Limited resources such as manpower, healthcare infrastructure, and funding present significant

challenges. In many cases, underfunded public health agencies are forced to prioritize certain tasks over others, leading to gaps in monitoring and underestimation of the pandemic's true extent.[82]

The presence of asymptomatic individuals complicates both estimation and monitoring. These individuals may unknowingly spread the disease, making it difficult to identify and contain clusters of infections.[83] Traditional surveillance methods often miss asymptomatic cases, leading to an underestimation of the pandemic's scope. Monitoring often necessitates data collection that may invade personal privacy.[84] Ethical considerations arise, especially when utilizing technologies like contact tracing apps, which may conflict with individual privacy rights. Striking a balance between public health and ethical imperatives is a complex challenge.

The estimation and monitoring of pandemics are fraught with challenges that have far-reaching implications for public health.[85] These challenges are multifaceted, spanning data inconsistency, public compliance, resource constraints, asymptomatic cases, and ethical concerns. Addressing these issues necessitates a multi-disciplinary approach that involves technological innovation, public awareness campaigns, and international collaboration. As pandemics continue to threaten global health, understanding and overcoming these challenges is of paramount importance.[86]

1.1.2. Approaches

1.1.2.1. Models for Estimation

In the realm of public health, the application of various epidemiological models has been integral for predicting and managing the spread of infectious diseases, particularly during pandemics. Established models like the Susceptible-Infected-Recovered (SIR) and Susceptible-Exposed-Infected-Recovered (SEIR)

compartmental models have played foundational roles, albeit with occasional criticism for their simplistic assumptions about population dynamics.[87]

Examining historical epidemics and pandemics, such as the Spanish Flu of 1918, SARS in 2002-2003, MERS in 2012, and H1N1 in 2009, researchers have employed a range of these models.[88] These past events serve as valuable case studies, shedding light on the effectiveness of different modelling approaches. For instance, compartmental models proved effective in providing general estimates during the H1N1 pandemic, while the complexity of transmission dynamics in SARS and MERS necessitated the use of agent-based models.[89]

Comparative analysis suggests that while traditional models are useful for generating general estimates, they may fall short in addressing the complexities and heterogeneities of real-world scenarios.[90] Advanced models, while offering finer details, come with the requirement for substantial computational resources and data.[91] Thus, a balanced approach that leverages the strengths of various models could provide the most accurate and actionable insights for public health authorities.

Despite the plethora of models available for estimating epidemiological trends in pandemics, a lack of standardization and a clear understanding of their respective merits and limitations persist. An examination of their application in historical pandemics serves as a valuable source of lessons, aiding in the refinement of existing models and the development of new ones that are better suited to the intricacies of modern epidemics.[92]

1.1.2.2. Models for Monitoring and Forecast

Forecasting teams leverage diverse models for monitoring and predicting various aspects of pandemics, incorporating data such as COVID-19 statistics, mobility trends, and demographic information.[93] Among the existing models, compartmental models like Susceptible-Infected-Recovered (SIR) and Susceptible-Exposed-Infected-Recovered (SEIR) serve as foundational tools, offering general estimates but criticized for simplistic assumptions.[94] Agent-based models, which consider individual interactions, provide greater granularity in understanding complex transmission dynamics but require substantial computational resources and data.[95] These models often use diverse data sources and estimations of intervention impact to forecast metrics like the number of cases, fatalities, and hospitalizations. Comparing and contrasting these diverse projections becomes crucial for discerning the efficacy of different models throughout the epidemic. This process involves accumulating and analyzing data, including identifying median points that serve as general or overall forecasts.[96]

Forecasting techniques play a vital role in designing effective plans and making informed decisions during pandemics. These methods rely on historical data to generate more accurate predictions about future events, aiding in the preparation for prospective threats and understanding their implications.[97] The precision of forecasts is contingent on data obtained from various platforms, encompassing environmental factors, incubation time, quarantine effects, age, gender, and other relevant variables. Despite their importance, forecasting techniques face unique challenges, both generic and technical, underscoring the necessity for a balanced and comprehensive approach in pandemic management.[98]

1.1.2.2.1. Ideal Statistical Model

Pandemics unfold as intricate phenomena shaped by numerous factors, encompassing biology, social dynamics, public policy, and healthcare infrastructure.[99] Constructing statistical models to project the trajectory of a pandemic necessitates a comprehensive, multi-dimensional approach. Nevertheless, prevailing models often simplify these complexities for computational convenience, potentially compromising accuracy and adaptability.[100]

An optimal statistical model should exhibit a high level of granularity, capable of capturing both population-wide trends and nuanced individual-level dynamics. This involves integrating variables such as age, gender, geographical location, and pre-existing medical conditions, acknowledging their significant impact on disease susceptibility and transmission.[101] Given the dynamic nature of pandemics, an ideal model demonstrates flexibility, enabling the incorporation of new data inputs and real-time adjustments to parameters. This adaptability is crucial for precise short-term forecasting, facilitating immediate public health interventions.[102] Flexibility also implies the model's ability to accommodate different diseases, strains, or variants without necessitating foundational restructuring.

The consideration of socio-economic variables is imperative for an effective model, encompassing factors like public healthcare infrastructure, social mobility, and compliance with public health measures.[103] Neglecting these variables could lead to misleading projections and ineffective policy recommendations. Additionally, the utility of a model is enhanced when it integrates real-time data streams, incorporating case numbers, vaccination rates, and mobility data.[104] Real-time analytics provide actionable insights for policymakers and healthcare providers, supporting timely and informed decision-making.

The ideal statistical model for pandemic forecasting amalgamates key elements: granularity, flexibility, consideration of socio-economic variables, and real-time data integration.[105] Adhering to these principles, such a model holds the potential to provide more accurate projections and contribute to effective, targeted interventions, thereby mitigating the impact of pandemics on global health and well-being.[106] Importantly, it is acknowledged that no single model comprehensively addresses all aspects, emphasizing the importance of selecting the most fitting or best models for specific applications.

1.1.2.2.2. Model Efficacy

Numerous statistical models exhibit significant disparities in their attempts to forecast the future impacts of the viral infection, utilizing diverse criteria such as social interaction distancing measures imposed in specific regions. While these models are invaluable tools for researchers worldwide, particularly when uncovering non-apparent facts, they often omit crucial information.[107] Additionally, epidemic models fall short of addressing economic challenges. Although many statistical models demonstrate mathematical soundness in their forecasts, it is prudent to apply a foundational level of computation across all models and scenarios, allowing for the inclusion of human-influenced variables like social distancing.[108] This approach enables the identification of the most fitting model for a given situation, potentially informing public use in future planning.

The article titled "Wrong but Useful — What COVID-19 Epidemiologic Models Can and Cannot Tell Us" by Inga Holmdahl and Caroline Buckee delves into the role and limitations of epidemiologic models within the context of the COVID-19 pandemic. The authors posit that while these models serve as essential planning tools

for policymakers, clinicians, and public health practitioners, they are frequently misconstrued or misapplied.[109] Despite their inherent inaccuracies in predictions, these models guide decision-making by presenting a spectrum of possible outcomes based on varying scenarios and assumptions. Regarding model efficiency, the paper underscores that simpler models may sometimes be more useful for immediate decision-making than their more complex counterparts.[110] Simpler models demand fewer computational resources and are more easily interpretable, rendering them highly efficient for rapid policy decisions. However, the authors caution that these models have their limitations and should not be perceived as definitive predictors of future events.[111]

The study also confronts the challenge of model uncertainty, asserting that all models inherently possess inaccuracies to some degree but still be useful. The efficiency of a model, therefore, should not be solely assessed based on its predictive accuracy but also on its adaptability to new data and evolving conditions. The authors advocate for iterative approaches that enable models to be updated as additional data becomes available, thereby enhancing their utility and efficiency over time. The study offers a nuanced perspective on the use and efficiency of epidemiologic models in managing the COVID-19 pandemic. It advocates for the appropriate application of models as planning tools, highlights the trade-off between simplicity and complexity in model efficiency, and encourages an iterative approach to model enhancement. This balanced viewpoint contributes to an understanding of how models are both flawed and useful, providing an effective guide for stakeholders involved in the pandemic response.[109]

1.2 Literature Review

The introduction to COVID-19 and statistical modelling provides an overview of the virus's origin and global spread, while also emphasizing the importance of employing statistical models to forecast its trajectory. Kai et al. (2020) examined the origins and worldwide dissemination of COVID-19, shedding light on the urgency of implementing robust statistical models to predict its future course. As the virus rapidly spread across the globe, understanding its behavior and forecasting its impact became imperative to guide public health interventions and policy decisions.[112] In a related study, Almeida et al. (2023) emphasized the value of data-driven approaches in comprehending the complexities of pandemic scenarios, with COVID-19 serving as a prominent example. Utilising statistical models that are informed by real-world data allows researchers and authorities to make informed decisions in managing the pandemic. These models assist in predicting infection rates, identifying potential hotspots, and evaluating the effectiveness of various control measures.[113]

Data processing and clustering methods are essential components in handling and analyzing pandemic datasets, particularly in the context of COVID-19. Sarker et al. (2022) emphasized the significance of data pre-processing techniques to manage missing and noisy data commonly found in real-world pandemic datasets. Data pre-processing involves cleaning, transforming, and organizing the data to ensure its quality and reliability. In the case of COVID-19 data, there may be instances of missing data or data points affected by noise, which impact the accuracy of analysis and modelling. By employing appropriate data pre-processing techniques, researchers mitigate these challenges and ensure that the data used for analysis is more accurate and reliable, leading to more robust insights and conclusions.[114]

In addition to data pre-processing, Li et al. (2020) introduced novel clustering methods designed specifically for grouping COVID-19 cases based on common characteristics. Clustering is a technique used in statistical methods to identify similar groups within a dataset. In the context of COVID-19, clustering methods were applied to identify patterns and trends among different cases, such as grouping cases with similar symptoms, transmission routes, or geographic locations. By organizing COVID-19 cases into clusters, researchers and decision-makers gain a better understanding of the virus's behavior and tailor specific strategies and interventions to different clusters. This aids in more effective decision-making and resource allocation, as well as informing targeted public health measures to control the spread of the virus.[115]

Time-series analysis and forecasting are essential techniques in understanding the temporal patterns and predicting trends in COVID-19 data. Liu et al. (2019) highlighted the significance of time-series theory, which involves analyzing data collected over time to identify patterns, trends, and seasonality. Time-series data in the context of COVID-19 would include daily or weekly case counts, hospitalizations, or mortality rates over a specific period. Autoregressive models, such as ARIMA (Auto Regressive Integrated Moving Average), are powerful tools commonly used in time-series analysis to forecast future values based on past observations. These models take into account the data's autoregressive nature, where the current value depends on past values and any relevant seasonal patterns or trends. By applying time-series theory and ARIMA models to COVID-19 data, researchers make predictions trajectory of the pandemic, enabling public health authorities to better plan and allocate resources.[116]

Gupta and Patel (2020) conducted a study applying ARIMA models to forecast COVID-19 cases in different countries, including India, the USA, and Italy. ARIMA models are particularly useful in capturing the temporal patterns and variations in COVID-19 data, making them suitable for predicting the number of cases over time. By utilizing historical data on COVID-19 cases, the models identify trends and patterns and generate forecasts for the future number of cases. This information is invaluable for policymakers and healthcare officials in preparing for potential surges in cases, anticipating resource needs, and formulating appropriate response strategies. The accuracy of these forecasts depends on the quality and completeness of the input data, making data integrity and pre-processing critical factors in obtaining reliable predictions.[117]

Control strategies and optimal allocation of resources are crucial in managing and mitigating the spread of the COVID-19 pandemic. Thompson et al. (2022) examined different control strategies that were employed to control the transmission of the virus. Trigger control involves implementing specific actions or interventions when certain predefined triggers or thresholds are met. These triggers were based on various factors such as the number of cases, the reproduction number (R_0), or the capacity of healthcare facilities. Optimal control theory is another approach that uses mathematical models to determine the most effective combination of interventions to minimize the impact of the pandemic. By optimizing control measures such as social distancing, testing, contact tracing, and quarantine, authorities effectively curb the spread of the virus.[118]

Lo-Ciganic et al. (2022) focused on the application of reinforcement learning algorithms for resource allocation in healthcare systems during pandemics. In the

context of COVID-19, these algorithms are utilized to optimize the allocation of limited resources such as hospital beds, ventilators, and medical supplies. By continuously learning from the outcomes of different resource allocation strategies, reinforcement learning algorithms adapt and improve their decision-making process over time. This helps in efficiently allocating resources where they are most needed, especially during the surge in cases, to ensure optimal care and outcomes for patients.[119]

Faiq (2020) investigated various methods for estimating Case Fatality Rate (CFR) in infectious diseases, including the use of Yoshikura's method. CFR is a crucial epidemiological indicator that measures the proportion of deaths among confirmed cases of a disease. It helps in assessing the severity and lethality of the disease outbreak. Yoshikura's method is a statistical approach for estimating CFR that takes into account the time delay between case confirmation and death, which was particularly important during a pandemic where there were delays in reporting and recording deaths. Accurate estimation of CFR is crucial for public health authorities to gauge the impact of the disease and formulate appropriate response strategies.[120]

Liu et al. (2021) employed distribution fitting techniques, specifically Weibull and Gamma distributions, to model COVID-19 mortality rates. Distribution fitting involves finding the best-fitting statistical distribution that represents the observed data. Weibull and Gamma distributions are commonly used to model survival data, including time-to-event data, such as the time from infection to death in the case of COVID-19. By fitting these distributions to mortality data, researchers better understand the underlying patterns and trends in mortality rates and make more informed predictions course of the pandemic. This information is valuable for

healthcare planning, resource allocation, and policymaking during a public health crisis.[121]

State-space methods are powerful tools for modelling and analyzing epidemiological data during disease outbreaks, providing real-time insights and predictions to guide public health responses. Afzal et al. (2022) discussed state-space methods for modelling epidemiological data. State-space models are a class of statistical models that capture both the observed data (e.g., the reported number of COVID-19 cases) and the underlying unobserved or hidden states (e.g., the true number of infected individuals). These models are particularly useful during outbreaks when there are delays in reporting cases or underreporting. State-space methods enable researchers to estimate the true state of the outbreak and provide more accurate and real-time insights into the disease's dynamics, including the transmission rate, the number of undetected cases, and the impact of control measures. These insights are crucial for public health officials to make informed decisions and implement timely interventions.[122]

Sharma et al. (2023) presented a case study of state-space estimation methods applied to predict COVID-19 transmission in specific states in India - Kerala, Karnataka, Tamil Nadu, and Maharashtra. State-space estimation allows researchers to assimilate various data sources, including reported case counts, testing rates, and demographic information, to estimate the true state of the outbreak and forecast its future trajectory. The study aimed to assess the effectiveness of different control measures and understand the varying dynamics of COVID-19 in different regions. By utilizing state-space methods, the researchers could gain insights into the pandemic's

progress, assess the impact of containment measures, and make recommendations for optimizing public health responses in each state.[123]

Oakley et al. (2020) conducted an assessment of the demand and capacity of hospital infrastructure, specifically focusing on beds, ICUs, and ventilators, to cope with the increasing number of COVID-19 cases. Understanding the healthcare system's capacity and potential limitations is crucial during a pandemic, as a surge in cases quickly overwhelms medical facilities. The researchers used data-driven methods and statistical modelling techniques to predict the demand for healthcare resources based on the projected increase in COVID-19 cases. By analyzing the available capacity of hospitals and comparing it with the estimated demand, the study aimed to identify potential gaps and areas of concern in healthcare preparedness. The findings provided valuable insights for policymakers and healthcare administrators to optimize resource allocation and plan for potential surges in COVID-19 cases.[124]

Bennett et al. (2021) investigated the prediction of the time it would take for available healthcare facilities to reach their maximum capacity during the COVID-19 pandemic. This assessment is crucial for anticipating and preparing for the potential overload of hospitals and critical care units. The study employed various statistical and mathematical models, along with real-time data on the pandemic's progression, to estimate the point at which healthcare facilities would be stretched to their limits. Understanding this critical threshold helps in planning and implementing timely interventions to prevent overwhelming the healthcare system and ensuring that adequate resources are available to provide the best possible care to COVID-19 patients.[125]

Both studies (Oakley et al. (2020), and Bennett et al. (2021)) contributed important insights into healthcare infrastructure planning during the COVID-19 pandemic.[124,125] By analysing demand and capacity, these research efforts aid in better understanding the potential strain on the healthcare system and support evidence-based decision-making for resource allocation and preparedness. They play a vital role in helping healthcare facilities and policymakers adapt to the evolving demands of the pandemic and ensure that healthcare services are efficiently delivered to those in need.

1.3 Justification for the study

The significance of this study lies in its comprehensive exploration and analysis of various statistical models, data processing techniques, and control strategies related to the COVID-19 pandemic. By investigating these aspects, the research aims to provide valuable insights and understanding of the pandemic's behaviour, transmission patterns, and impact on healthcare systems. The study covers a wide range of topics, using data reconciliation, time-series analysis, clustering methods, and predictive modelling.

The study's findings contribute to the development of more effective and accurate statistical models for predicting the spread of the virus, identifying relevant variables, and assessing the effectiveness of containment measures. This is instrumental in guiding policymakers and health authorities in making informed decisions and adopting targeted interventions to manage and control the pandemic. Moreover, the analysis of different epidemiological models, such as SIR and SIRD models, provides a deeper understanding of the dynamics of the virus's transmission and the potential implications on the population's health.[126] Additionally, the study

delves into the case fatality rate, recovery rate, and distribution fitting, aiding in assessing the severity of the disease and estimating mortality rates.

The significance of this study extends to its application in specific regions, such as health decision-making in health and pollution control. By analysing the data and trends in these areas, the research provides valuable insights into the regional variations in COVID-19 cases and the adequacy of healthcare infrastructure. This study's multifaceted analysis of statistical models and data-driven methods serves as a valuable resource for researchers, policymakers, and healthcare professionals seeking to gain a deeper understanding of the COVID-19 pandemic and implement evidence-based strategies to combat its spread and mitigate its impact on public health and the healthcare system.

1.3.1 Existing Models

While existing statistical models have proven instrumental in understanding and forecasting the dynamics of the pandemic, they are not without their limitations. Many of these models rely heavily on historical data and assumptions that do not hold as the pandemic evolves. The accuracy of predictions is hindered by factors such as changing public behaviour, the emergence of new virus variants, and variations in testing and reporting practices across regions. Additionally, existing models struggle to account for complex interactions between different interventions and the interplay of social, economic, and healthcare factors. It is imperative to acknowledge these limitations to refine and advance statistical modelling approaches for more effective pandemic management.[127]

Furthermore, the reliance on historical data poses challenges in addressing the unprecedented nature of the current pandemic. The evolving dynamics of COVID-19, characterized by emerging variants and shifting public responses, necessitate continuous adaptation of statistical models.[128] The limitations become apparent when dealing with unforeseen variables that were not present in the historical data, leading to potential inaccuracies in forecasting future trends. As the scientific community grapples with these challenges, it underscores the importance of not only acknowledging the limitations but also actively working towards enhancing the resilience and adaptability of statistical models in the face of dynamic and complex public health crises. This recognition serves as a catalyst for ongoing research and refinement, driving the evolution of statistical modelling to better meet the demands of pandemic management.[129]

1.3.2 Need for Real-Time Monitoring

In the face of a rapidly evolving pandemic, the need for real-time monitoring and data-driven decision-making has never been more pronounced.[130] Traditional epidemiological models often rely on retrospective data, resulting in a lag between data collection and actionable insights.[131] Real-time monitoring systems, underpinned by advanced statistical models, are crucial for the timely detection of outbreaks, assessment of intervention effectiveness, and resource allocation.[132] These systems empower public health authorities with the agility to adapt strategies in response to changing circumstances and to mitigate the pandemic's impact effectively.[133] This study explores the development and application of such real-time monitoring approaches in the context of COVID-19, addressing the pressing need for dynamic and data-informed pandemic management.

The integration of data-driven techniques further enhances the capabilities of real-time monitoring systems.[134] By leveraging these advanced technologies, the models were not only detecting patterns and anomalies in data more efficiently but also adapting and learning from evolving trends. This synergy allows for more accurate predictions, identification of emerging hotspots, and a deeper understanding of the factors influencing the spread of the virus.[135] As the pandemic landscape continues to evolve, the exploration of innovative technologies within the realm of real-time monitoring becomes pivotal. This study delves into the intersection of advanced statistical models, and real-time monitoring, aiming to contribute insights into the development of adaptive and data-driven strategies for effective pandemic management. Through an exploration of these interconnected elements, the research endeavours to provide a comprehensive understanding of the potential for cutting-edge technologies to revolutionize the landscape of public health response during dynamic and challenging times.

1.4 Research Questions

1. How does the established model for the COVID-19 pandemic function, and what are its key components?
2. What are the specific methods and strategies used in monitoring the COVID-19 pandemic?
3. How does the use of the established model for the COVID-19 pandemic help the decision-making process of planners and healthcare providers in managing and containing the pandemic?
4. What are the strengths and limitations of the established model for the COVID-19 pandemic compared to traditional approaches in monitoring and managing the pandemic?
5. What is the comparative analysis of pandemic management outcomes between regions that have adopted the established model?
6. How adaptable is the established model for the COVID-19 pandemic to future pandemics or infectious disease outbreaks?

1.5 Objectives of the Study

- To establish the model for the COVID-19 pandemic.
- To provide methods for monitoring of COVID-19 pandemic, so that planners and healthcare providers can use the models for appropriate action to manage, contain, and control the pandemic.

CHAPTER 2: MATERIALS AND METHODS

2.1 Introduction

The study focus is given to data-driven schemes, which are methodologies that predominantly utilize empirical data for analysis, eschewing reliance on purely theoretical or heuristic models.[135] Data-driven schemes refer to techniques that rely heavily on the empirical data collected, rather than purely theoretical or heuristic models.[136] Amidst the COVID-19 situation, these schemes play a crucial role in providing immediate and accurate information about the transmission, severity, and trends of the illness.[137] Data-driven techniques are necessary because of the unpredictable nature of the COVID-19 pandemic.[138] Traditional models, which relied on historical data or pre-set parameters, are not agile enough to capture the nuances and rapid changes associated with the virus's spread.[139] Data-driven schemes, on the other hand, adapt and evolve as more data becomes available, ensuring that predictions and estimations remain relevant and accurate.[140]

2.1.1. Holistic Approach

The COVID-19 crisis has introduced trade-offs between economic stability, public health, and social well-being, and policymakers find themselves grappling with complex decisions.[141] As the pandemic spread during the initial months of 2020, it swiftly morphed into a global epidemic, casting a shadow over the majority of the world's population. With stringent containment measures in place, an unprecedented lockdown affecting vast populations ensued, marking an unprecedented chapter in human history.[142]

Without a doubt, the COVID-19 epidemic has caused the most substantial economic decline since World War II, in addition to its catastrophic impact on human health. Disruption reverberated across all sectors of the economy, as global supply chains faltered, demand for imported goods diminished, international tourism dwindled, and business travel ground to a halt.[143] Governments have found solace in analytical techniques and modelling to decipher, strategies, predict, and respond to the multifaceted implications of this crisis.[144] These techniques encompass a spectrum of decision support tools and statistical methodologies, often harnessed through multi-agent models. These models encompass disease surveillance, predicting epidemic outcomes, and counterfactual studies, underpinned by a diverse array of data-driven techniques.[145]

Data-driven tools play a crucial role in the context of pandemic outbreaks. They are essential for (i) modeling and predicting the course of the pandemic, (ii) implementing timely interventions to control and stop the spread of the infection, and (iii) monitoring the evolution of the epidemic and evaluating the potential impacts of different healthcare and socioeconomic measures.[146] The process of making effective decisions during a pandemic is complex and filled with uncertainty. Concurrently, the inability to promptly and efficiently react in the face of immense uncertainty results in severe repercussions.

Transitioning to the realm of a data-driven community, which encompasses researchers and practitioners designing decision-making tools and predictive models, a holistic approach becomes pivotal.[147] In summation, the holistic approach of combining data-driven tools with a SWOT analysis empowers decision-makers to navigate the intricate landscape of challenges. By understanding the internal dynamics

and external influences, practitioners develop targeted strategies, enhancing their ability to tackle prevailing issues, identify favourable conditions, and overcome hurdles.[148] This approach stands as a beacon of precision and insight, guiding scientific endeavours toward effective solutions in the face of multifaceted challenges.

2.1.2. Data-Driven Schemes and Techniques

Governments around the globe are increasingly relying on various analytical techniques and modelling to comprehend, plan, forecast, and respond to the pandemic.[149] These techniques include decision support techniques and statistical methods using multi-agent models, such as: (i) disease surveillance and (ii) predicting epidemic outcomes (such as mortality, hospital demands, and case counts), that used a diverse range of data-driven techniques, such as Bayesian techniques, time-sequencing ARIMA models, and (iii) counter-factual study of epidemics, and the results seem to have a significant impact on the initial decisions regarding lockdowns in different nations.[150]

Data-driven tools are critical in the context of epidemic outbreaks to (i) model and predict the trajectory of the pandemic; (ii) control the infection by taking prompt action to reduce and stop the contagion.; and (iii) track the epidemic's progression and evaluate the potential effects of any countermeasures taken from both a healthcare and a socioeconomic perspective. Within the framework of a pandemic, optimal decision-making is a sophisticated procedure that involves a large level of ambiguity; whereas, concurrently, failing to react quickly and appropriately, even in the face of overwhelming uncertainties, that have serious implications.[151]

This research included the elements of data-driven schemes in the following manner. (i) Data Collection: This entails acquiring pertinent data from diverse sources, open-source platforms, Government testing centers, and public health organizations. The quality and granularity of this data are crucial for the success of the subsequent analysis.[152] (ii) Data Processing: After gathering, the data undergo cleaning, standardisation, and transformation to a format that is appropriate for analysis. This stage often entails addressing the presence of missing numbers, outliers, and other biases within the data.[153] (iii) Data Analysis: With the processed data in hand, various statistical techniques are applied to derive insights, make predictions, and inform decision-making processes. By leveraging real-time data and advanced analytical techniques, these schemes provide invaluable insights that inform public health decisions, policy-making, and community responses. As the pandemic continues to evolve, the role of data-driven schemes in estimation and monitoring will remain central to global efforts to understand, mitigate, and eventually overcome the challenges posed by COVID-19.[154]

2.2 Materials

2.2.1. Data Sources

Open data resources are crucial in combating the COVID-19 pandemic. The scientific community conducts a daily analysis of time series data on the number of confirmed cases and mortality rates, along with other indications, and publishes it on open source.[155] It is crucial to depend on precise, up-to-date, and varied data sources and the main data sources used in this research are:

- **World Health Organization (WHO) Data Repository**: The World Health Organisation (WHO) issues daily status bulletins about the COVID-19 pandemic, which include comprehensive information on the number of cases, fatalities, and recoveries categorised by country. This repository serves as a central store of worldwide data, including both total statistics and daily additions. [156]
- **Johns Hopkins University (JHU) COVID-19 Dashboard**: The Centre for Systems Science and Engineering (CSSE) at JHU maintains a frequently updated dashboard that monitors the worldwide dissemination of the virus. This site provides detailed data, including statistics at the province or state level in several nations. [157]
- **National Health Departments**: For a more detailed country-specific analysis, data from national health departments and ministries were extracted. These sources often provide insights into local outbreaks, testing rates, and hospitalization figures.[158]
- **COVID-19 Data Repository by the Centre for Humanitarian Data**: This repository offers a comprehensive dataset that includes testing rates, case counts, and mortality figures. It also provides demographic data, which is crucial for understanding the disease's impact on different population groups.[159]
- **Scientific Literature and Preprints**: Peer-reviewed articles and preprints from platforms like medRxiv and bioRxiv were consulted to gather insights on the virus's transmission dynamics, clinical characteristics, and potential interventions.[160]

2.2.2. Overview of Open Source COVID-19 Data

Compilation of open source repositories with relevant COVID-19 data includes:

- **Confirmed/Active cases, Death or Recovered cases**: Data sets are used to monitor the count of deaths, verified new cases, and instances of recovery. Data is often provided on a national level and, in some cases, on a regional level. Up until this point, Johns Hopkins University (JHU) has maintained the most popular data set. Additionally, there are regional repositories for every nation. The consequences of the pandemic on human life are broadly represented by these types of data sets.[161]
- **Testing**: Details on the number of positive incidents that happened, the type of test conducted, and the number of tests are included in the datasets connected to proactive testing. It is also crucial to have accessibility to auxiliary data, such as the gender and age categories, professional activities, and the methods used to choose the test subjects.[162]

2.2.3. Auxiliary Data Sources

- **Demographics**: Standardization of the COVID-19 data is done while considering demographic data to construct general models and procedures. The factor that could help to explain why the virus spread so quickly in some areas is population density. Additionally, age groupings are used in estimating COVID-19 mortality.[163]
- **Hospital and Healthcare Systems**: The data that pertain to healthcare systems cover a lot of different factors, one of which is the number of ventilators and intensive care units (ICUs). The national healthcare systems are responsible for maintaining the data about the availability of healthcare resources, which, in certain data sets, is only partially available.[164]

2.2.4. Data Curation

Integration and organization of data gathered from multiple sources is known as data curation. The availability, preservation, and reuse of data are critical aspects of effective data management, particularly in the context of disease outbreaks like COVID-19.[165] Data curation, is an active process, that involves organizing and maintaining datasets to meet analytical requirements and preferences. Obtaining datasets is just the initial step; the goal of data curation is to make them easily discoverable, comprehensible, and accessible. Metadata management, including the use of data catalogues, plays a crucial role in data curation.

During outbreaks, extensive data are collected globally, but challenges in data sharing arise due to legal, ethical, and privacy concerns. The lack of standardized formats hinders efficient sharing. Establishing a standard format is essential, accommodating various data streams from reputable sources.

Rapid data ingestion requires automated workflows and these Integrated process flows enhance data collection comprehension, crucial during early outbreak stages when data quality varies. Scalability concerns were addressed through a decentralized approach to data validation involving team members and volunteers. Using a standardized geographic reference frame enables detailed and integrated data analysis. The data for the study was collected automatically from various sources and securely stored in designated arrays, as indicated in the metadata.

2.3 Methods

2.3.1. Basic Methods

2.3.1.1. Pre-Processing Data

Due to the heterogeneous origin of data, the majority of real-world data are very prone to missing, noisy, and inconsistent. Data analysis methods would not produce high-quality results when applied to this noisy data because they would be unable to successfully find patterns. Therefore, data pre-processing is crucial to raising the overall level of data quality.[166]

Data pre-processing is a crucial step in statistical analysis, encompassing procedures necessary for encoding or modifying data to facilitate efficient decoding. The precision and accuracy of a model's predictions heavily rely on the algorithm's capacity to rapidly analyze the data's properties. In statistical terms, the effectiveness of a predictive model is contingent upon the careful pre-processing of data, ensuring that it meets the criteria for accurate statistical analysis and robust modelling. This involves encoding or transforming data to enhance its suitability for statistical algorithms, enabling a thorough examination of its features and patterns. The success of statistical analyses and subsequent modelling hinges on the meticulous preparation of data through pre-processing procedures, ultimately influencing the reliability and validity of the statistical outcomes.[167]

Missing or duplicate values could result in an inaccurate interpretation of the data's overall statistics. Various strategies are employed to address these gaps:

- Data Imputations: This involves replacing missing values with estimated ones. The objective is to create a complete dataset. Techniques range from

simple mean imputations to more complex methods like k-nearest neighbours or multiple imputations.

- Data Transformations: At times, the data is transformed to a different scale or format to make it more suitable for analysis. For instance, logarithmic transformations are used to address skewness in data distributions.
- Dealing with Duplicates: Duplicate values, if unnoticed, lead to redundancy and bias. They are typically identified and removed to maintain the integrity of the research.
- Use of Robust Statistical Methods: Some research opts for statistical techniques that are less sensitive to missing or outlier values, ensuring that the overall findings remain credible.
- Exclusion of Missing Data: In some cases, it is more appropriate to omit data points with missing values, especially if their inclusion could distort the findings or if they represent a negligible portion of the dataset.

Contradictory data points and outliers frequently tend to interfere with the model's efficiency, producing inaccurate predictions.[168]

2.3.1.2. Data Reconciliation

In the context of statistical analysis, potential errors in data proceedings and mapping logic during migration introduced uncertainties and inaccuracies in the dataset. These errors, ranging from inconsistencies in encoding to issues with mapping algorithms, have significant implications for the reliability of statistical analyses. In statistical terms, these issues lead to skewed distributions, inaccurate measures of central tendency, and compromised statistical inferences. Addressing and rectifying these challenges in data migration is essential to ensure the integrity of the

dataset and, consequently, the validity of statistical analyses conducted on the migrated data.[169]

Data reconciliation (DR) is a data verification method that takes place during data migration. In this step, it is determined whether the migration is successfully transmitting data by comparing the target data to the source data. Since the accessible data sources have several severe restrictions, data reconciliation techniques are crucial in the proposed approach.[115] To create time series with improved quality and identify abnormalities in the actual data, methodologies including signal processing, data fusion, data reconciliation, and clustering were utilized. It is easy to get carried away in the mapping and transformation logic throughout the data migration procedure. Data is left in an incorrect condition as a result of several types of errors. These lead to a variety of problems, including records that have gone missing, values that are missing, incorrect values, records that have been duplicated, values that have been badly formatted, or systems that are broken.[117] Hence it is very important to use data reconciliation techniques.[170]

These issues result in a variety of challenges, including:

- Improperly formatted values
- Missed records
- Erroneous values
- Missed values
- Duplication of records

Without any of the data reconciliation steps, these problems are undetected, severely impacting the accuracy of the entire data and producing false. When it comes to data reconciliation, the conventional method frequently relies on straightforward record counts to determine whether the anticipated number of records were transferred. Usually, this was because field-by-field verification took a significant amount of processing power. The problem with this is missing records, just one of them, during the migration process. Consequently, the other problems won't be noticed.[171]

Data reconciliation helps to extract reliable and precise information on the state of the industrial process from the measurement data. It also helps you create a single data set that indicates the most likely process operation. This results in inaccurate information. The various types of data reconciliation techniques include:

- **Master data reconciliation:** The master data is reconciled between source and target in master data reconciliation. Mostly, the master data changes slowly or static, and no aggregation technique is applied to the dataset. [172]
- **Transactional data reconciliation:** The total sum is utilized in the transactional data reconciliation procedure to avoid any mismatches created by modifying the granularity of qualifying dimensions.[173]
- **Automated data reconciliation:** By making data reconciliation an inherent part of data loading, it is simple to automate the process of data reconciliation in data warehouse management systems. It enables you to maintain distinct tables for importing metadata.[174]

2.3.1.3. Data Fusion

The process of combining data from various sources to create information that is more reliable, accurate, and helpful than data from any single source alone. Depending on the step of processing at which fusion occurs, data fusion procedures are frequently characterized as high, intermediate, and low. Data fusion is the collaborative application of knowledge from several sources to help understand phenomena as a whole. Data fusion algorithms were categorized as phenomenological and non-phenomenological. The method for fusing data is derived from phenomenological algorithms employing an understanding of the physical processes as a foundation. These strategies are being pursued by several researchers. Such techniques are probably difficult to develop and time-consuming to use.[175]

Contrarily, non-phenomenological techniques frequently neglect the physical process and aim to combine information by employing statistics related to individual data points. In this context, it is essential to create efficient data fusion methods capable of utilising such multi-sensor features, to properly utilise the data collected by several sensors or by the same sensor in various measured conditions. The goal of data fusion research is to combine the sub-information from several sensors to create a comprehensive understanding of the system under analysis. [176]

2.3.1.4. Clustering Methods

In the context of COVID-19 analysis, cluster analysis involves the identification of similar groups of objects based on relevant data points. The goal is to categorize elements into distinct clusters, where all members within a cluster share common characteristics. This process allows for the organization of COVID-19 data

into meaningful categories or clusters. For instance, one cluster encompassed data points related to symptomatic patients, while another cluster could include data points of asymptomatic individuals. The primary objective of cluster analysis in this study is to discern patterns and relationships within COVID-19 data, facilitating a more nuanced understanding of different manifestations of the disease.[177]

2.3.1.5. Real-Time Surveillance

This is done by proactive testing, tracing, monitoring mobility, etc. The implementation and development of surveillance systems' capability of detecting second pandemic waves are equally relevant in this context. It was used to detect events such as monitoring the impact on the health care system and estimating event magnitude by applying new case definitions, describing the natural history of disease and establishing case definitions for COVID-19 as well as tracking the change in disease severity, evaluating event response by assessing the changes in the incidence and monitoring.[178]

The combat over COVID-19 calls for modelling techniques to play a significant role. Epidemiological models were from the complex spatially distributed models to low dimensional compartmental. The modified models provide fundamental factors that define the virus's capacity to spread. Additionally, data-driven parametric inference offers tools for foreseeing the outcomes of chosen interventions. However, due to significant issues like non-linearities, incomplete observations, and non-identifiability, certain strategies are needed to fit the models to the existing data. It is necessary to implement model evaluation, sensitivity analysis, validation, and selection procedures. There are other options besides the forecasting options provided by epidemiological models. In this situation, a variety of data

science forecasting methods were used. Simple linear parametric techniques are options. The approaches are either non-parametric or parametric. A few of these techniques give probability descriptions of the predictions made. However real-time surveillance of the outbreak is an essential component of the 3M approach.[179]

2.3.1.6. Time-Series Theory

Several advanced approaches from the fields of stochastic processes, time series analysis, and signal processing are used in the analysis of COVID-19. When describing the COVID-19 raw daily prevalence data, for example, they were utilized to enhance the reliability of the initial time series. A statistical technique called time series analysis works with data from time series or trend analysis. Time series data refers to information that is organized into a series of discrete time intervals or periods.[180] The data falls into three categories:

- **Time series data**: A collection of observations about the values a variable takes at different times.
- **Cross-sectional data**: Data from one or even more variables that have been collected simultaneously.
- **Pooled data**: A mix of cross-sectional and time series data.
- **Dependence**: The relationship between two observations made by the same factor at prior time points is known as dependence.
- **Stationarity**: Demonstrates the average value of a series that doesn't change over time; stationarity is not achieved if historical influences accumulate and values rise to infinity.

- **Differencing**: Used to manage the autocorrelations, to make the series stationary, and to de-trend; Certain time series analyses do not require over-differenced and differencing series, that produce erroneous estimations.

- **Specification**: Evaluating the non-linear or linear relationships of dependent variables using models including ARIMA.

- **Time series analysis and exponential smoothing**: Based on the past and present value, this approach forecasts the value for one following period. It includes averaging the data in a way that each case or observation's non-systematic components cancel each other out. The short-term prediction is made using the exponential smoothing method. Parameters used to estimate the impact of the time series data are delta, phi, alpha, and gamma. Alpha is applied if there is no evidence of seasonality in the data. Gamma is used whenever a sequence of data demonstrates a trend. Delta is used if the data shows seasonality cycles. A model is used following the data's pattern.[181]

2.3.1.7. The 3M Concept

Making optimal decisions within the framework of COVID-19 is a complicated procedure that necessitates dealing with a substantial level of uncertainty as well as the severe repercussions of not responding promptly and with sufficient intensity. A road map is provided that leads from the data source towards the final decision-making phase. Monitoring, Modelling, and Making decisions are the three components of the proposed 3M-an3M analysis

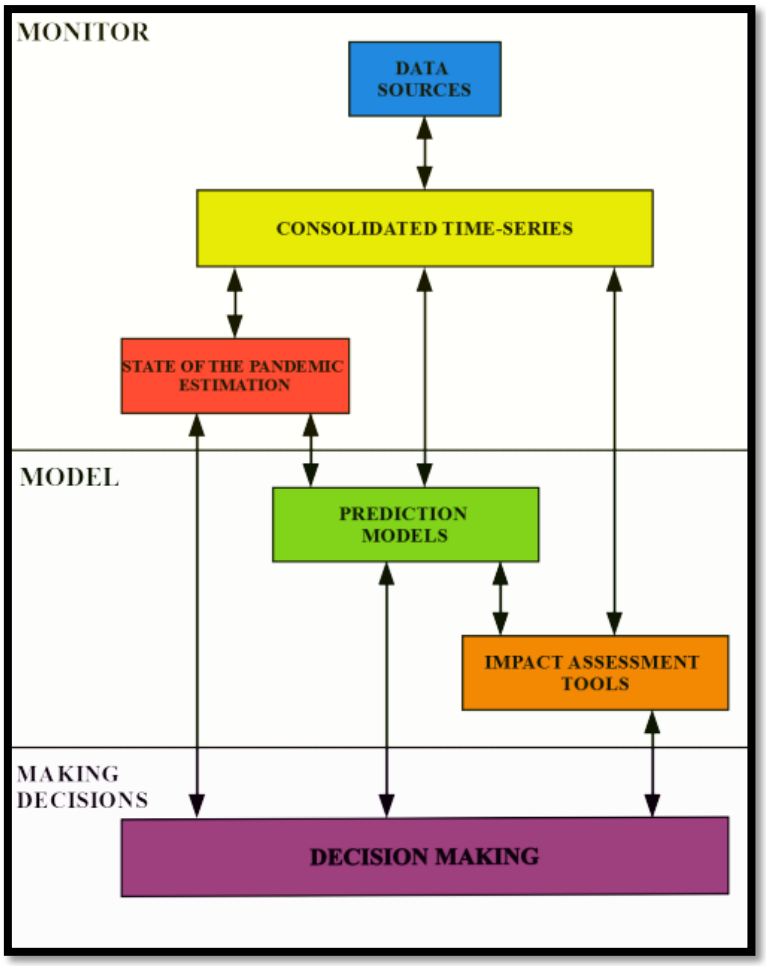


Figure 1: Workflow Diagram for Pandemic Data Analysis and Decision Support System

2.3.2. Monitoring

Monitoring in the context of COVID-19 involves the systematic collection, analysis, and interpretation of health data essential for planning, implementing, and evaluating public health practices. An effective monitoring system is crucial to manage the spread of the virus, assess the effectiveness of interventions, and support decision-making processes. During the pandemic, monitoring has taken on an unprecedented role, with real-time data being used to track infection rates, hospitalizations, recoveries, and fatalities. This has enabled health authorities to identify hotspots, allocate resources, and inform the public about the risks and

necessary precautions. The integration of various data sources into a consolidation series has provided a more comprehensive view of the pandemic's trajectory, assisting in predicting future trends and potential impacts. By leveraging data-driven methods, public health officials monitor the spread of COVID-19 with greater precision, leading to more informed strategies for containment and mitigation.[183]

2.3.2.1 SWOT Analysis of Data-Driven Methods in the COVID-19 Pandemic

The SWOT (Strengths, Weaknesses, Opportunities, and Threats) framework was used to evaluate the COVID data analytical model. The SWOT analysis is a strategic planning tool that helps identify and assess the internal and external factors that affect success. COVID-19 SWOT analysis is essential for developing a communication rhythm and effective decision-making skills in challenging situations. The goal of a COVID modelling SWOT is to classify the immediate negative and positive impact of external and internal forces, resulting in a four-category SWOT matrix.[184]

Strengths: The strengths of statistical models in the context of COVID-19 are numerous. They offer a predictive lens through which we estimate the trajectory of the virus, helping to anticipate case surges and enabling proactive responses. These models synthesize complex and varied datasets into actionable insights, which are critical for informing public health strategies. By evaluating trends and patterns, statistical models have been fundamental in forecasting demand for medical supplies, hospital capacity needs, and the potential impact of public health interventions. This modelling has been indispensable for strategic planning, allowing governments and health organizations to prepare and respond effectively to the evolving pandemic.[185]

Weakness: Analyzing the weaknesses of statistical models of COVID-19 is of paramount importance in the ongoing battle against the pandemic. Statistical models have played a crucial role in informing public health decisions, resource allocation, and policy implementation. However, they are not without limitations. Identifying and addressing these weaknesses is essential for improving the accuracy and reliability of COVID-19 predictions.[186]

Opportunity: Analyzing the opportunities presented by statistical models of COVID-19 is crucial for several reasons. These models play a significant role in shaping public health policies and strategies for managing the pandemic. Understanding the opportunities, helps policymakers make informed decisions to mitigate the spread of the virus, allocate resources efficiently, and protect vulnerable populations.[187]

Threats: The threats associated with relying on statistical models for COVID-19 management stem primarily from the uncertainty and variability of pandemic dynamics. Models are inherently built on assumptions that do not always capture the rapidly changing conditions or the emergence of new virus variants. There is also the risk of data inaccuracy or incompleteness, which leads to misguided predictions and responses. Public misinterpretation of model predictions further complicates the public health response, leading to complacency or panic. Moreover, overreliance on models overshadows the need for adaptable and flexible policy-making that quickly responds to new information beyond what models predict.[188]

2.3.2.2 Consolidated Time-Series

Consolidated time-series data refers to the aggregation of time-stamped data points into a coherent, chronological sequence. In the context of COVID-19, this data includes daily or weekly counts of cases, hospitalizations, recoveries, deaths, and other relevant information, recorded over a specific time frame. These data points are crucial for understanding the dynamics of the pandemic and making informed decisions.[189]

Consolidated time series in COVID-19 estimation and monitoring are done for the following:

- Trend analysis: Consolidated time-series data allows statisticians and epidemiologists to identify trends and patterns in the spread of the virus. This is essential for predicting future outbreaks and understanding the effectiveness of interventions.
- Model calibration: Statistical models, such as SEIR (Susceptible-Exposed-Infectious-Removed), rely on consolidated time-series data for calibration. By fitting the model to historical data, researchers fine-tune parameters and make more accurate predictions.
- Epidemiological parameters: Parameters like the reproduction number (R_0) and the serial interval are estimated using time-series data. These parameters are critical for assessing the virus's transmission dynamics and the impact of control measures.
- Resource allocation: Healthcare systems rely on consolidated time-series data to allocate resources effectively. Hospitals anticipate surges in cases and plan for the allocation of beds, ventilators, and medical staff.

- Evaluation of interventions: Monitoring the effects of interventions, such as lockdowns, mask mandates, and other campaigns, is possible through time-series data. It helps policymakers understand which measures are most effective in curbing the spread of the virus.
- Early warning systems: Time-series data is used to develop early warning systems that alert authorities to potential outbreaks or hotspots, allowing for swift responses to contain the virus's spread.

Using the consolidated time-series data, several statistical models were developed:

- ARIMA (Auto Regressive Integrated Moving Average): This model captures the autocorrelations in the time-series data and has been used to forecast future cases.
- SEIR (Susceptible, Exposed, Infected, Recovered) model: This compartmental model provides insights into the spread of the virus in populations over time.

Consolidated time-series data forms the backbone of statistical models used in the estimation and monitoring of COVID-19. It provides a rich source of information that enables accurate predictions, effective resource allocation, and evidence-based decision-making. As the pandemic continues to evolve, the importance of high-quality, timely, and consolidated data cannot be overstated in the battle against COVID-19.[190]

a. Clustering Methods

Clustering analysis is applied to time-series COVID data to identify patterns or trends over time. By clustering data points from similar periods, the study monitors

and categorizes phases or waves of the pandemic, understanding the similarities or differences in disease progression, transmission rates, or other relevant metrics. This helps in real-time monitoring and forecasting potential future trends based on past clustered patterns. Cluster analysis is a multivariate method that aims to classify subjects based on a set of measured variables into several different groups such that similar subjects are placed in the same group. These methods group the data sets into clusters, such that the data sets within one cluster are more related to each other than the data sets in different clusters. In the context of COVID-19, these methods help identify patterns in the spread of the virus by grouping regions, individuals, or cases with similar epidemiological and demographic characteristics. [191]

There are several clustering techniques commonly applied to COVID-19 data:

- K-means clustering: K-means is a widely used clustering method that partitions data into K clusters. In COVID-19 analysis, K-means group areas with similar infection rates, population density, or healthcare infrastructure, helping identify hotspots or areas with unique characteristics.
- Hierarchical clustering: Hierarchical clustering creates a tree-like structure of clusters, making it useful for exploring data at multiple scales. In the context of COVID-19, this method revealed relationships between regions, allowing for the identification of clusters of interest.
- Agglomerative clustering: Agglomerative clustering starts with individual data points and progressively merges them into larger clusters. In monitoring COVID-19, this method highlighted how cases or regions evolve over time, revealing emerging trends.

Hierarchical cluster analysis was used in this study for the estimation and better lockdown relaxation strategies that could be implemented are also suggested in the study. To form a recommendation cluster, the data of COVID-19 cases among different states in India are used.[192] Common methods include:

Euclidean Distance

$$d(A,B) = \sqrt{\sum_{i=1}^{n}(A_i - B_i)^2}$$

Manhattan Distance:

$$d(A,B) = \sum_{i=1}^{n}|A_i - B_i|$$

The study examines the shifting pattern of the COVID-19 outbreak across various states in India before and during the lockdown. This analysis is based on hierarchical clustering of states using data on confirmed, death, and recovered cases. Two distinct periods are considered: the 20 days leading up to the lockdown (March 5, 2020 - March 24, 2020) and the 20 days of the lockdown (March 25, 2020 - April 13, 2020). The baseline for cluster analysis comprises Indian states with more than 500 confirmed cases.

To account for the fact that people were stranded in different states due to the lockdown, the study uses the total population of each state as the population residing in that state during the lockdown period. Additionally, hierarchical cluster analysis is applied to the trend line coefficients of the respective states, with the resulting dendrograms illustrating the shift in trends. The study incorporates data on the number of individuals in different Indian states who were confirmed COVID-19 cases, those who succumbed to the disease, and those who recovered from it.[193]

- **Quartile Based Clusters**

A quartile divides the data set into three points, a lower quartile Q1 (25%), median Q2 (50%), and upper quartile Q3 (75%). Now, we map out the four groups formed from these quartiles. The first group contains the lowest number up to Q1; the second includes Q1 to the median; the third contains median to Q3; the fourth category contains Q3 to the largest value in the data set.[194]

The method of locating quartiles is similar to that method used for finding the median. Q1 is the value of the item at $(n + 1)/4^{th}$ position and Q3 is the value of the item at $3(n + 1) / 4^{th}$ position when actual values are known.

For a row data if there are odd numbers of observations there will be only one middle value and it will be the median. That means if there are n observations arranged in order of their magnitude, the size of (n+1)/2nd observation will be the median or Q2. If there are even numbers of observations the average of two middle values will be the median. Median will be the average of $n/2^{th}$ and $(n/2) +1^{th}$ observation.[195]

b. Time-Series Theory

A Box-Jenkins approach to time series analysis is used for forecasting the confirmed cases of COVID-19 in India. A time series is a sequential set of data points, measured typically over successive times. It is mathematically defined as a set of vectors x(t), t = 0,1,2,... where t represents the time elapsed. The variable x(t) is treated as a random variable. The measurements taken during an event in a time series are arranged in a proper chronological order. It was used to examine how the changes

associated with the chosen data point compare to shifts in other variables over the same period.[196]

I. Time Series Plot

A time series plot of the data is sketched to evaluate the trend of the data. The time-series plot is the most frequently used form of graphic design. It is a 2-dimensional plot in which the X-axis shows the range of days and the Y-axis shows the COVID-19 cases corresponding to the respective days.

II. ARIMA Model

The ARIMA (p, d, q) model is given by:

$$\phi(B)(1-B)^{d}\ Z_t = \theta(B)a_t$$

where we define,

$$\phi(B) = 1-\phi_1 B-\phi_2 B^2-\phi_2 B^2-.....-\phi_p B^p$$

$$\theta(B)=1-\theta_1 B-\theta_2 B^2-\theta_2 B^2-.....-\theta_q B^q$$

Where B is the backshift operator, it is a white-noise process, ϕ and θ are the AR and MA model parameters respectively. The future value of a variable is assumed to be a linear–combination of past observations and a random error together with a constant term in an AR(p) model. The integration parameter d is a nonnegative integer. When d = 0 we have the usual ARMA model, that is ARIMA (p, 0, q) ≡ ARIMA(p, q).[197]

Mathematically the AR(p) model is expressed as:

$$y_t = c+\varphi_1 y_{t-1}+ \varphi_2 y_{t-2}+\ldots\ldots\ldots\ldots\ldots+ \varphi_p y_{t-p}+ \varepsilon_t$$

MA(q) model is given by:

$$y_t = \mu +\theta_1 \varepsilon_{t-1} +\theta_2 \varepsilon_{t-2} +..... +\theta_q \varepsilon_{t-q} + \varepsilon_t$$

Here y_t and ε_t are respectively the actual value and random error at period t, ϕ_i (i = 1,2,...,p) are model parameters and c is a constant. An ARMA (p, q) model is a combination of AR(p) and MA(q) models which are obtained from ACF and PACF plots.[198]

The autoregressive (AR)(p) time series model component is used for the forecast of future values of dependent observations. The moving average (MA)(q) model deals with past forecast errors for the forecast of the future dependent value. Both the AR and MA combination make the ARIMA model, to deal with the stationary data value. A third component, integrating (I) to convert the observations using a differencing series is utilized for the non-stationary values. The differencing order value was used for the model forecast for COVID-19 confirmed cases to avoid any misleading observed.[199]

III. Augmented Dicky Fuller Test

The stationarity of the data is scrutinized through an Augmented Dickey-Fuller (ADF) test. The Augmented Dickey-Fuller test allows for higher-order autoregressive processes by including Δy_{t-p} in the model. But our test is still if $\gamma = 0$.

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 y_{t-2} \ldots$$

The null hypothesis is, H_0: $\phi_1 \neq 1$ i.e., the process contains a unit root and it is non-stationary and the alternative hypothesis is, H_1: $\phi_1 = 1$ i.e., the process does not contain a unit root and therefore it is stationary. [200]

The level of significance (α) is taken as 0.05. If the p-value exceeds α , we accept the null hypothesis and if the p-value is less than 0.05, we reject the null hypothesis.

IV. Model Selection Criteria

a. Autocorrelation and Partial Autocorrelation Function

From ACF and PACF plots, we determine the order of AR(p) and MA(q). The amount of correlation between a variable and a lag of itself is a partial autocorrelation. The autocorrelation of a time series y at lag 1 is the coefficient of correlation between y_t and y_{t-1}, which is presumably also the correlation between y_{t-1} and y_{t-2}.[201]

b. Akaike Information Criteria

Akaike's information criterion compares the quality of a set of statistical models to each other.[202]

It is usually calculated as:

$$AIC = -2(\text{log-likelihood}) + 2K$$

Where,

- K is the number of model parameters.
- Log-likelihood is a measure of model fit. The higher the number, the better the fit. This is usually obtained from statistical output.

ARIMA (0, 1, 2) is identified as the better model with the lowest AIC value.

The Auto-Regressive Integrated Moving Average (ARIMA) model is fitted for the data to generate short-term predictions and to analyse the disease-based trajectory model. Time series data with trends or seasonality will affect the value of the data at different times.[203]

V. In-Sample Forecasting

In-sample forecasting is done to determine the performance of the ARIMA (0,1,2). It utilizes a subset of the available data to forecast values outside the estimation period. The in-sample forecasted data is cross-checked with real-time data and if not, much variation is observed then accepted.[204]

VI. Residual Analysis

Residual analysis is performed to check the model's adequacy. The basic assumption in residual analysis is that the variables are uncorrelated. The Conditional Sum of Squares (CSS) method is used for estimating the ARIMA (0,1,2) model in the residual analysis. The residuals in a time series model are what is left over after fitting a model. For many time series models, the difference between the observations and the corresponding fitted values gives the residuals.[205]

Residual = Estimated value - True value

$$e_t = y_t - \hat{y}_t$$

Where, y_t is the observed value of model parameters and $\hat{y}_t$ is the fitted value.

VII. Shapiro Wilk Test

The null hypothesis of the Shapiro-Wilk test is that the population is normally distributed. The test statistic is:

$$W = \frac{(\sum_{i=1}^{n} a_i x_i)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Where, W is the test statistic, n is the sample size, x_i is the ordered sample values and a_i are the constants generated from means, variances and covariance from the order statistics from a normal population and $\bar{x}$ is the sample mean. The level of significance is taken as 0.05. If the p-value < 0.05, the null hypothesis is rejected and concluded that the test data is not normally distributed.[206]

VIII. Ljung Box Test

The basic assumption in residual analysis is that the variables will be uncorrelated. The autocorrelation of the data is tested through the Ljung Box test. It is a diagnostic test used for the lack of fit of a time series model. The Ljung Box test is applied to the residuals of a time series after fitting an ARMA (p,q) model to the data. The test examines n autocorrelations of the residuals and if the values are small, the model exhibits no significant lack of fit. Here the null hypothesis, H_0: the model does not exhibit a lack of fit is tested against the alternative hypothesis, H_1: the model exhibits a lack of fit.[207]

The test statistic is given by

$$Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k}$$

Where Q is the test statistic, n is the sample size in the time series, $\hat{\rho}_k$ is the estimated autocorrelation at lag k, and h is the number of lags being tested. The test rejects the null hypothesis if $Q > \chi^2_{(1-\alpha,h)}$, where $\chi^2_{(1-\alpha,h)}$ is the Chi-square value with $h = m - p - q$ degrees of freedom and α level of significance (here p and q are the parameters of ARMA model fit to the data).

IX. Forecasting

The trajectory of COVID-19 is predicted through a short-term forecast. After an appropriate time, series model has been fit, the model is used to generate forecasts of future values. If the current time is denoted by T, the forecast for $y_{T+\tau}$ is called the τ-period-ahead forecast and is denoted by $\hat{y}_{T+\tau}$ (T). The accurate method to use in identifying the best forecast is the mean squared error for which $E[(y_{T+\tau} - \hat{y}_{T+\tau}(T))2]$

$= E[e_T(\tau)^2]$, is minimum.[208] It was shown that the best forecast in the mean square sense is the conditional expectation of $y_{T+\tau}$ given current and previous observations, that is,

$$y_T, y_{T-1}, \ldots : \hat{y}_{T+\tau}(T) = E[y_{T+\tau} | y_T, y_{T-1}, \ldots]$$

X. Interrupted Time Series Analysis

An Interrupted Time Series analysis (ITS) is used for tracking a period before and after a point of intervention to assess the lockdown effect. To identify the same, the data is classified into two groups, from January 30, 2020 to March 24, 2020 (before lockdown) and from March 25, 2020 to May 8, 2020 (during lockdown). The effect of an intervention on an outcome variable either for a single treatment group or when compared with one or more control groups was estimated using the ITS command. The options in ITS analysis allow control for auto-correlated disturbances and to estimate treatment effects over multiple periods.[209]

2.3.2.3 Estimation of the State of the Pandemic

a. Measures of Disease Severity

Disease severity among selected Indian states and union territories is estimated using Case Fatality Rate (CFR), Recovery Rate (RR), and Mortality Rate (MR). States and union territories in which at least 100 cases are reported as of 31st May 2020 are selected for measuring severity.

I. Case Fatality Rate

In epidemiology, CFR is used to measure the disease severity and to predict the disease outcome.[210] The CFR is a ratio, which gives the proportion of people who die from a particular disease out of all the infected people during a certain period. CFR ranges between 0 and 1 (0% and 100%); it measures the risk of a disease in terms of disease mortality. Higher CA higher indicates poor outcomes. In the case of COVID-19 disease, CFR acts as a better estimate for understanding the outbreak and epidemiological features of the disease.[211]

CFR is calculated by using the following formula:

$$CFR = \frac{Number\ of\ deaths\ due\ to\ a\ particular\ disease}{Total\ number\ of\ cases\ due\ to\ same\ disease} * 100$$

II. Recovery Rates

Recovery rate often refers to the rate of transition from a state of being infected with a disease to a state of recovery from that disease. Recovery rates are used for understanding disease dynamics. Recovery rates give us an understanding of the progress of an infectious disease in a group of people. In this study, we are using the recovery rates of selected South Indian states to estimate the disease severity.[212] Recovery rates calculated by using the formula:

$$RR = \frac{Number\ of\ people\ recovered}{Number\ of\ cases\ with\ outcome} * 100\ \%$$

III. Mortality Rate

Mortality rate also referred to as death rate is the measure of the number of deaths in a population to the total population per unit of time.[213] It gives the frequency of death in a defined population in a specific interval of time. It is expressed in units of death per 1000 individuals per year. The mortality rate is calculated by using the following formula:

$$MR = \frac{Deaths\ occurring\ during\ a\ given\ time\ period}{Size\ of\ the\ population\ among\ which\ the\ deaths\ occurred} * 10^n$$

In the case of the COVID-19 pandemic, the mortality rate is given by:

$$MR = \frac{\text{Number of deaths due to C0VID} - 19}{Total\ population} * 100000$$

It is the number of deaths from COVID-19 per 100,000-person year at risk. It is usually calculated separately for different geographic areas. In this study, we are estimating the mortality rate of COVID-19 disease for each selected state separately over a particular period.

b. Real-Time Epidemiology

The purpose of this method was to estimate the basic reproduction or the basic reproduction ratio (R_0) of COVID-19 in India using mathematical models and also to find the growth factor using the exponential growth model. The basic reproduction number is defined as the expected number of infections produced by a single infected person in a susceptible population.it is a measure of the potential spread of the disease in a population. When$R_0 > 1$ the infection is expected to continue in the population and if $R_0 < 1$ disease will die out. The dynamics of the infection depend on the basic reproduction ratio. To estimate the basic reproduction number, the study employs the

basic Susceptible-Infected-Removed (SIR) and Susceptible-Infected-Recovered-Death (SIRD) epidemic models.[214]

c. SIR model

The SIR model considered in this study is based on the simple case proposed by Kernmak and McKendrick in 1932. An SIR model is an epidemiological model that computes the theoretical number of individuals infected with a contagious illness in a closed population over time. In an SIR model, the whole population is divided into three compartments: Susceptible(S), Infected (I), and Removed(R) which include both recovered and deceased.

Susceptible denoted by S: Individuals in the population who have not been infected. They are healthy but at risk of becoming infected. Once they have contracted the infection, they move into the infected sub-group. Infected is denoted by I: Infected individuals who are contagious or are carriers. They infected susceptible individuals. Removed is denoted R: Individuals who have recovered or died from the disease. Unfortunately, the SIR model does not describe the difference between immunity, non-immunity, and even innate immunity.[215]

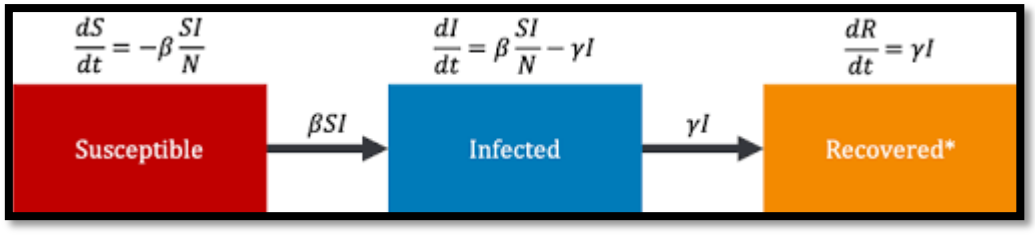


Figure 2: The diagrammatic representation of the SIR model

The basic assumptions of the SIR model[216]:

- The total size of the host population remains constant (S + I + R = N).
- The population should mix homogeneously.
- It will not allow any host demographic turnover (either birth or death) in the period of the epidemic, and all infections are assumed to end with recovery or removal from compartments.
- A person leaves or discharges from the susceptible compartment only by becoming infected.
- A person leaves or discharges from the infected compartment only by recovering from the disease.
- The probability of being infected does not depend on factors such as age, gender, or social status.
- During epidemics, susceptible individuals isolate themselves from infection, or take other protective measures.
- The recovery rate is constant over time.
- The model assumes that a recovered person will become immune and hence cannot spread the disease.
- The dynamical equations are of first order:

  dS /dt = − rate of new infections

  dI /dt = rate of new infections − rate of recovery

  dR /dt = rate of recovery

The following is the first-order differential equation:

$$\frac{dS}{dt} = -\beta \frac{SI}{N}$$

$$\frac{dI}{dt} = \beta \frac{SI}{N} - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

$$N = S + I + R$$

New infections occur through contact between infected and susceptible individuals, and the rate of change is proportional to the number of interactions. The number of susceptible individuals decreases as individuals come into contact with the infected:

$$\frac{dS}{dt} = -\beta \frac{SI}{N}$$

When susceptible become infected, they leave the susceptible compartment and join the infected. Thus, the total number of infected individuals increases. Vice versa, the hosts leave the infected compartment and join the recovered group. Since β is assumed constant, this implies that the rate of change is dependent on time as the size of the infected group varies:

$$\frac{dI}{dt} = \beta \frac{SI}{N} - \gamma I$$

Since infected individuals only leave the compartment by joining the new Removed compartment, it only changes through the addition of those recovered from infection. The recovery rate is given by the constant parameter.

$$\frac{dR}{dt} = \gamma I$$

The parameter controlling how often a susceptible-infected contact results in a new infection is the transmission rate β and the rate at which an infected person recovers and moves into the resistant phase is called the recovery rate γ .

To determine if there is an epidemic, we look at the steadiness of the disease-free equilibrium. We only need to consider the variable I (t). The condition for an epidemic to occur is > 0.

We have:

$$\beta \frac{SI}{N} - \gamma I > 0$$

$$\frac{\beta}{\gamma} \frac{S}{N} > 1$$

At the beginning of an epidemic, everyone in the population is susceptible, i.e. implies S ≈ 1. For S = 1, we obtain the condition

$$R_0 = \frac{\beta}{\gamma}$$

The dynamics of the infection depend on the basic reproduction ratio.

d. SIRD Model

The Susceptible-Infected-Recovered-Death (SIRD) model is an extension of the SIR model where the recovered compartment is split into recovered and deceased. In the SIRD model, the whole population N is subdivided into four compartments: Susceptible(S), Infected (I), Recovered (R), and Dead (D).[217]

The differential equations for the SIRD model:

$$\frac{dS}{dt} = -\beta \frac{SI}{N}$$

$$\frac{dI}{dt} = \beta \frac{SI}{N} - \gamma I - dI$$

$$\frac{dR}{dt} = \gamma I$$

$$\frac{dD}{dt} = dI$$

Where the parameters are:

β : transmission rate

γ : recovery rate

d : death rate

- It will not allow any host demographic turnover (either birth or death) in the period of the epidemic, and all infections are assumed to end with recovery or removal from compartments.
- A person leaves or discharges from the susceptible compartment only by becoming infected.
- A person leaves or discharges from the infected compartment only by recovering from the disease.
- The probability of being infected does not depend on factors such as age, gender or social status.
- During epidemics, susceptible individuals isolate themselves from infection, or take other protective measures.
- The recovery rate is constant over time.

- The model assumes that a recovered person will become immune and hence cannot spread the disease.

To find the basic reproduction number, the available data, and the predicted values were fitted using the least square method. The least-square fit chooses a parameter value x which minimizes the sum of the square of residuals. Curve fitting by the method of least squares concerns combining a set of measurements to derive **estimates** of the parameters that specify the curve that best fits the data.[218,219]

e. Exponential Growth Model

The exponential growth model gives the number of cases at a certain time point. The first period of epidemics follows exponential growth.[220]

The formula of exponential growth is:

$$x(t) = x_0 * b^t$$

Where $x(t)$ is the number of infections at t, x_0 initial number of cases and b number of people infected by each sick person (growth factor).

For estimating the growth factor the study needs to rewrite the exponential growth model formula in the form:

$$y = a + d * x$$

The study needs logarithms to rewrite the exponential growth formula in a form that has linear regression:

$$\log x(t) = \log(x_0) + \log(d) * t$$

The study performed linear regression analysis to estimate the values of constants (a and b). The study models the epidemic curve by employing exponential growth. The predicted values using the exponential growth model and the infected number of cases reported are fitted using linear regression to estimate the growth factor.

f. Epidemic wave Insights

A statistical model was developed to study the Epidemic wave and the duration of patients in COVID-19 for each state of India. Data was taken directly from the first day of reporting for each state.[220] Telangana was removed from the analysis due to the inappropriateness of the reported data. The complete data of Tripura was also unavailable in the dataset.[221]

A fundamental relation between Incidence and prevalence is given as:

$$\text{Prevalence odds} = (\text{Incidence}) * (\text{Duration})$$

$$\text{Expected number of diseased individuals}(\text{Number of new cases per time unit}) * (\text{Expected duration of disease})$$

$$\text{Expected duration of disease} \frac{\text{Expected number of diseased individuals}}{\text{Number of new cases per time unit}}$$

The regression analysis has been used to understand the average duration of COVID-19 in various Indian states. This study used the cumulative number of new cases as the predictor variable and cumulative active cases as the outcome variable. Log transformations were used to attain the assumptions of the regression model. The first case reporting day in each state was used as the linear regression model's starting point. The coefficient of determination (R^2) is evaluated to assess the good fit of the

model. The 95% confidence interval (CI) is calculated from the standard error of the slope.

The models to estimate the Active Cases on i^{th} day was:

$$\hat{Y} = Ae^{a_1 t + a_2 t^2} \text{ and } \hat{Y} = Be^{b_1 t + b_2 t^2 + b_3 t^3}$$

Where, A= e^{a_0}, B = e^{b_0} and $a_0, a_1, a_2,$ and b_0, b_1, b_2 and b_3 are the Regression Coefficients.

95% Confidence Interval (CI) of predicted active cases on each day was computed as:

$$95\%\ \text{CI}$$

$$= e^{\text{Predicted Number of log(Active Cases)} \pm 1.96 * \text{SE(Predicted number of log(Active Cases))}}$$

$$95\%\ \text{CI} = e^{Ae^{a_1 t + a_2 t^2} \pm 1.96 * \text{sqrt}(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d})}$$

Where, S.E. for Quadratic model was computed as:

$$\text{S. E. (Predicted Number of active cases at } x_i \text{ day)}$$

$$= \text{S. D.} \sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} + \frac{(x_i^2 - \bar{x^2})^2}{\sum (x_i^2 - \bar{x}^2)^2}}$$

Where S.D. is the Standard Deviation of y, which is calculated as:

$$\text{S. D.} = \sqrt{\frac{\sum (y - \hat{y})^2}{n - 1}}$$

Where, y is the log transformation of seven days moving average of Total Active Cases, $\hat{y}$ is the predicted number of the Active Cases, and 'n' is the number of days Active Cases were observed.

2.3.3. Modelling

2.3.3.1 Growth Models

The mathematical growth models develop a differential equation, using the dynamic nature of the disease and the rate of its spread to better predict future predictions.[222]

Assumptions:

- The population is closed.
- Lockdown restrictions imposed on the population remain the same throughout the projection period.
- The deaths and recoveries among the diseased population are not considered.
- Age–dependent effects in the transmission of COVID-19 are not considered.

a. Exponential Growth Model

The exponential model is associated with the name of Thomas Robert Malthus (1766-1834) who first realized that according to a geometric series, any species could potentially increase in numbers. The growth rate of the population in exponential growth is directly proportional to the size of that population. In the caseCOVID-19-19, exponential growth at a particular point in time gives the number of cases (number of infections). Several epidemiological studies suggest that exponential growth happens in the first period of any epidemic.[223]

The mathematical model for exponential growth is:

$$I(t) = I_0 \mathrm{e}^{\mathrm{rt}}$$

Where *I(t)* is the number of infections at time *t. I_0* is the initial number of cases, *r* is the number of people infected by each sick person *(*growth factor), which is an important measure for the spread.

Exponential growth just fits the epidemic in its initial stages, there are situations like recovered people no longer spread disease or every individual in the population becomes infected, and then we have to use some other growth model that will consider situations like hitting the carrying capacity of growth. [224]

b. Logistic Growth Model

The logistic growth model or Verhulst model, after the name of Belgian mathematician Pierre Verhulst, is used to describe biological systems in connection with population growth among different restrictions (limited resources for growing). It is commonly used to explore the risk factors of a certain disease, and to predict the likelihood that a certain disease will occur according to the risk factors. Through logistic regression analysis, the development and transmission law of epidemiology were predicted. At first, the logistic growth model is roughly exponential, but it has a reduced growth rate as the output approaches the upper bound of the model, called the carrying capacity. The model of logistic growth is a mathematical model applicable in many situations.[225]

In logistic growth models, the growth rate increases at the beginning period, and then, as it reaches its maximum it begins to decline in the later stages. The COVID-19 outbreak was tracked using a logistic growth model. In the case of COVID- 19

disease this maximum limit refers to the total population in the world, because, when the whole world gets sick, the growth will necessarily lower. In short, logistic growth involves initially exponential growth, and then declines. The reason for using logistic growth to model the COVID-19 outbreak is that during the first period of its outbreak, these types of epidemics follow exponential increases.[226]

Logistic growth is expressed as:

$$I(t) = k/(1 + I_0 e^{-rt})$$

Where *I(t)* is the total number of infections at time t. I_0 is the initial number of cases, *k* is the carrying capacity of an epidemic and *r* is the maximum number of cases.

c. Gompertz Growth Model

The model was originally proposed as an animal population growth model by Benjamin Gompertz (Gompertz, 1825), to describe the population's extinction law.[160] Infectious disease development is similar to the growth of individuals and populations. The Gompertz model is selected in this study to describe the spread law of infectious diseases, and to study the factors that control and affect COVID-19 spread. COVID-19 growth resembles Gompertz's growth at the beginning and end of time when there is slow growth. This model gives the best fit when the dataset follows a smooth curve.[227]

Gompertz's growth is given by;

$$I(t) = I_0 e^{-ke^{-rt}}$$

Where I_0 is the initial number of cases, k is the carrying capacity of an epidemic, r is the growth rate.

d. Bertalanffy Growth Model

The Bertalanffy model often serves as a model of growth. It is mainly used to study the factors which control growth and influence it.[161] It is used to describe the characteristics of fish growth. Also, other species were used to describe animal growth, such as pigs, horses, cattle, sheep, etc., and other infectious diseases. Infectious disease development is similar to the growth of individuals and populations. The Bertalanffy model is selected in this study to describe the spread law of infectious diseases and to study the factors that control and affect COVID-19 spread.[162] The spread law of the pandemic COVID-19 and related factors were studied using the growth-affecting Bertalanffy model. It deeply explains the factors affecting growth.[228]

The Bertalanffy growth model is;

$$I(t) = k(1 - e^{-rt})^{I_0}$$

Where I_0 is the initial number of cases, k is the carrying capacity of an epidemic, r is the growth rate.

a. Estimation of Parameters

- **Curve Fitting Method**

For the estimation of model parameters, the curve fitting method is adopted. Curve fitting is the way we model or represent a spread of data by assigning a 'best fit' (curve) function along the entire range. Ideally, it will capture the data trend, and

enable us to predict how the data series will behave in the future.[163] For a given set of data, curve fitting is the process of introducing mathematical relationships between dependent and independent variables in the form of an equation. For curve fitting the least square method is used.[229]

- **Method of Least Squares**

In fitting a curve by the method of least squares, the parameters of the curve are estimated by solving the normal equations which are obtained by applying the principle of least squares concerning all the parameters associated to the curve jointly (simultaneously). However, for a curve of higher degree polynomial and or for a curve having many parameters, the calculation involved in the solution of the normal equations becomes more complicated as the number of normal equations then becomes larger. Moreover, in many situations, it is not possible to obtain normal equations by applying the principle of least squares concerning all the parameters jointly. These lead to thinking of searching for some other method of estimating the parameters. For this reason, a new method of fitting a curve has been framed which is based on the application of the principle of least squares separately for each of the parameters associated with the curve. The method of least squares helps us to find the values of unknown parameters in such a way that the following two conditions are satisfied.[230]

The sum of the residual of observed values Y and corresponding expected values of Y will be zero,

$$\sum (Y - \widehat{Y}) = 0$$

The sum of the squares of the residual of observed values Y and corresponding expected values ($\widehat{Y}$) should be at least $\sum(Y - \widehat{Y})^2$.[230]

- **Model Evaluation**

Model Evaluation is an integral part of the model development process. It helps to find the best model that represents our data and how well the chosen model will work in the future. Evaluating model performance with the data used for training is not acceptable in data science because it easily generates overoptimistic and overfitted models. The efficiency of an estimator was measured by its performance. The fitting ability of growth models is compared using the coefficient of determination. The coefficient of determination is the square of the correlation (r). It measures the variability between the predictor variable and the variable expected or the deviation between the fitted line and the real data. It is also known as "goodness of fit". R^2 ranges from 0 to 1 (0% to 100%)**.** If the value ofR^2 is close to 1, it indicates that the model is a good fit for the data available. The variation increases as it deviates from 1. A value near to or equal to 0 shows no relation between the predictor and the predicted variable.[231] It was calculated by using the following formula;

$$R^2 = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}$$

Where TSS is the total sum of squares and RSS is the residual sum of squares i.e. the sum of squares due to regression**.**

2.3.3.2 Prediction Model

a. Distribution fitting of COVID-19

Here a statistical distribution is evaluated that best fits the data of COVID-19 mortality in India by using the MLE method. For distribution fitting, the choice of distribution is limited to continuous distributions. By inspecting the nature of our data, the candidate distributions are chosen. The major steps involved in distribution fitting are: the choice of suitable distributions, parameter estimation, and finding the best fit.[232]

The first step is the choice of candidate distributions for fitting distributions to the data. Henceforth histogram and empirical distribution were plotted to check the normality and afterward, the data was obtained as right-skewed. Then the Weibull, Gamma and Lognormal distributions for the data are considered. The next step is to estimate the value of the unknown parameter of the chosen distribution. The maximum likelihood estimation method (MLE) is the most commonly used method to estimate the value of unknown parameters. In this method, the likelihood of the function is maximized with respect to the parameter. To understand the best fit among the distributions, consider three values such as log likelihood, AIC, and BIC values are considered. A maximum log likelihood value and lower AIC and BIC value is the criterion for finding the best fit. The goodness of fit plots such as density plots, CDF plots, Q-Q plots, and P-P plots are plotted. These plots were also helpful in determining the best-fit distribution.[233]

To describe a distribution among a set of parametric distributions, descriptive statistics was used. Weibull, Gamma, and Lognormal distributions to find the best-

fitted distribution among them. f(.|θ) (with parameter $\theta \in R^{d}$) was fitted to the data set, one at a time, after selecting one or more parametric distributions.

I. Weibull Distribution

The probability density function of a Weibull random variable is:

$$f(X = x; \lambda,k) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{\left(-\frac{x}{\lambda}\right)^{k}} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

where k > 0 is the shape parameter and λ >0 is the scale parameter of the distribution

The Weibull distribution is a continuous probability distribution used to analyse life data, model failure times, and access product reliability. It is an extreme value of probability distribution that is frequently used to model reliability, survival, wind speeds and other data. The only reason to use Weibull distribution is because of its flexibility. Because it simulated various distributions like normal and exponential distributions. Weibull distribution reliability is measured with the help of parameters.[234] The two versions of the Weibull probability density function (pdf) are

- Two parameter pdf
- Three parameter pdf

II. Gamma Distribution

The gamma distribution represents continuous probability distributions of two-parameter families.[235] Gamma distributions are devised with generally three kinds of parameter combinations.

If X is a continuous random variable, then the probability density function is:

$$f(x|\alpha,\lambda) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \quad \text{with } x \in R^{+}$$

- α = the shape parameter.
- (sometimes θ is used instead) = The rate parameter (the reciprocal of the scale parameter).

α and λ are both greater than 1.

When α = 1, this becomes the exponential distribution.

When λ = 1 this becomes the standard gamma distribution.

Alpha and lambda define the shape of the graph.

III. Lognormal Distribution

The lognormal distribution is commonly used to model the lives of units whose failure modes are of a fatigue-stress nature. Since this includes most, if not all, mechanical systems, the lognormal distribution has widespread application. Consequently, the lognormal distribution is a good companion to the Weibull distribution when attempting to model these types of units. As surmised by the name, the lognormal distribution has certain similarities to the normal distribution. A random variable is log-normally distributed if the logarithm of the random variable is normally distributed. Because of this, there are many mathematical similarities between the two distributions. For example, the mathematical reasoning for the construction of the probability plotting scales and the bias of parameter estimators is very similar for these two distributions.[236] A positive random variable X is log-normally distributed if the logarithm of X is normally distributed

$$Ln(X) \sim N(\mu,\sigma^2)$$

Distribution parameters were by default estimated by maximizing the likelihood function.

a. MLE Method

The maximum likelihood estimator (MLE) of a parameter θ (or a series of parameters) was used as an estimate of the parameters of a distribution. For a sample the likelihood function is defined by $L(\theta) = \prod_{i=1}^{n} f(xi\,|\theta)$ according to the i.i.d. sample assumption, where x_i is the n observations of variable X and $f(.|\theta)$ is the density function of the parametric distribution.[237]

Goodness of fit

The goodness of fit test is a statistical hypothesis test to see how well sample data fit a distribution from a population with a normal distribution.[168] This test shows if the sample data represents what one would expect to find in the actual population or if it is somehow skewed. Goodness-of-fit establishes the discrepancy between the observed values and those that would be expected of the model in a normal distribution case.[169] Maximum goodness of fit estimation also called the minimum distance estimation method consists of three different kinds of distances. These are Cramer-Von Mises distance, Kolmogorov - Smirnov distance and the classical Anderson Darling distance.[238]

Along with the goodness of fit tests, the distribution graphs were very helpful in determining the best-fitting model. The fundamental difference of this approach is that it is quite subjective: while the goodness of fit tests is "exact" in the sense that the results do not depend on the researcher (provided that the tests are performed correctly), using various graphs is a more empirical way to analyze your data.

There are several commonly used tests, all of which tell whether a particular distribution is a good fit. However, these tests differ in how they are performed, and

sometimes do not agree with one another. For example, one finds out that according to the Kolmogorov- Smirnov test, the Weibull distribution is the best fit, but the Anderson-Darling test suggests that it is not. That is when the graphs come in handy.[239]

Four classical goodness-of-fit plots such as density plot, CDF plot, Q-Q plot and P-P plot are plotted.

- **Probability Density Function (PDF) Graph**

The Probability Density Function graph displays the theoretical pdf of the fitted distribution (or several distributions) and the histogram of the given sample data.

- **Cumulative Distribution Function (CDF) Graph**

The Cumulative Distribution Function graph displays the theoretical CDF of the fitted distributions and the empirical CDF based on the given sample data. While the PDF graph mainly shows the shape of the data given, the CDF graph is useful to determine how well the distributions fit to data[240]

- **Probability-Probability (P-P) Plot**

The probability-probability plot is a graph of the empirical CDF values plotted against the theoretical (fitted) CDF values. It is used to determine how well a specific distribution fits the observed data. The P-P plot will be approximately linear if the specified theoretical distribution is the correct model.[241]

- **Quantile-Quantile (Q-Q) Plot**

The quantile-quantile plot is a graph of the input data values plotted against the quantiles (inverse CDF values) of the fitted distribution.

These plots helped to determine the best-fitted distribution. Here we adopted the Kolmogorov-Smirnov statistic to identify the best fit model for the data. A maximum log-likelihood value and lower AIC and BIC value is the criterion for finding the best-fitted distribution. In this study, we compared the AIC values of the three distributions, and the distribution with the lowest AIC value was selected as the best-fitted model for the data. [242]

b. Akaike Information Criteria

The Akaike Information Criterion allows us to test how well the model fits the dataset without overfitting it.[243] The accuracy and performance of the model are measured using AIC and is given (approximately) by:

$$AIC = - 2 \ln (max. likelihood) + 2 p$$

Where p denotes the number of independent parameters estimated in the model. Thus, the AIC essentially chooses a model with the best fit, as measured by the likelihood function, subject to a penalty term that increases with the number of parameters fitted in the model. This should prevent overfitting.

c. Time-Series Modelling

Time-series modelling is used to obtain the trend and forecasting pattern of mortality of infection in India due to COVID-19 using Time-series modelling. ARIMA modelling is one of the best modelling techniques for forecasting a time series. It explicitly provides a set of standard structures in time series data, and as such provides a simple yet powerful method for skillful forecasts of time series.[244]

d. Data Preparation

Box-Jenkins model mainly involves model identification of the ARIMA process, model estimation or fitting and diagnostic checking. Then the fitted model is used for forecasting the time series data. Primarily, a time series plot is considered to examine the trend, and outliers of the data. It is usually obtained by plotting the observation against time. Here, the study acquires the time pattern changes after a specific point, and if so, we examine it further to find out the nature of change and potential causes. Box Jenkins modelling primarily includes the evaluation of the stationarity of the data using the Augmented Dickey-Fuller test. The existence of trend and periodicity will result in non-stationarity. If data exhibit a trend, we remove the deterministic trend or difference in the data. To remove these effects, the differencing stage is included in Box-Jenkins. Differencing is continued until the time series becomes stationary. Then we plot ACF and PACF which are useful in identifying and modelling patterns in time series. When the modelling is completed, the result will be summarized or integrated to produce the estimations and forecast.[245]

e. Autocorrelation

The stationary assumption allows us to make simple statements about the correlation between two successive values, Xt and Xt+k . This correlation is called the autocorrelation of lag k of the series. The autocorrelation function displays the autocorrelation on the vertical axis for successive values of k on the horizontal axis.[246]

Let X(t1), X(t2)…, X(tn) be a time series observation for a set of times t1, t2…,tn, then autocorrelation of order h is defined as;

$$\rho(h) = \frac{Cov(X_t, X_{t+h})}{\sqrt{Var(X_t)}\sqrt{Var(X_{t+h})}}$$

This formula giving the autocorrelation coefficients is known as the Autocorrelation Function (ACF) After the stationarity of the series is specified, the next task in Box-Jenkins analysis is to identify an appropriate model from the sample autocorrelation function.

f. Augmented Dickey-Fuller Test

In ARIMA (p,d,q), the trend and seasonality of the data are removed using a differencing method. The differencing is repeated until the series becomes stationary. The frequently performed method in autoregressive models for testing the stationarity of a time series is the Augmented Dickey-Fuller Test.

Another point to remember is the ADF test is fundamentally a statistical significance test.[247] That means, there is hypothesis testing involved with a null and alternate hypothesis and as a result, a test statistic is computed and the p-value gets reported. It is from the test statistic and the p-value, that an inference as to whether a given series is stationary or not was determined. Here the level of significance is taken as 0.05. The hypothesis is given as

H_0 : The data is not stationary

H_1 : The data is stationary

If the p-value is less than 0.5, we reject our H_0 and conclude that the data is stationary. Otherwise, data is non-stationary and has again undergone differencing.

g. Model Identification

The main aspect of time series analysis and modelling is the consideration of autocorrelation. ARMA models combine autocorrelation methods (AR) and moving averages (MA) with time-dependent parameters. When differencing is included in the

procedure, it becomes ARIMA (p,d,q) or Box-Jenkins modelling where 'p' stands for the order of auto-regression, 'd' signifies the degree of trend difference, and 'q' is the order of moving average.[248]

h. AR(p) Process

The p-th order autoregressive (AR) model — that is, AR(p) model is

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \ldots + \phi_p y_{t-p} + \epsilon_t$$

where $\{y_t\}$ is the data of daily COVID-19 deaths in India on which the ARMA model is to be applied. The parameters φ_1, φ_2, and so on are AR coefficients. The value of p is determined from the partial autocorrelations of the appropriately differenced series. If the partial autocorrelations cut off after a few lags, the last lag with a large value would be the estimated value of p. If the partial autocorrelations do not cut off, you either have a moving average model (p=0) or an ARIMA model with positive p and q.[249]

i. MA(*Q*) Process

The *q*-th order moving average (MA) model — that is, MA(*q*) model:

$$y_t = c + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \ldots - \theta_q \epsilon_{t-q}$$

Where $\{y_t\}$ is daily COVID-19 death data and θ_1, θ_2, and so on are MA coefficients. The value of q is found from the autocorrelations of the appropriately differenced series. If the autocorrelations cut off after a few lags, the last lag with a large value would be the estimated value of q. If the autocorrelations do not cut off, you either have an autoregressive model (q=0) or an ARIMA model with a positive p and q.[250]

j. ARMA(*P,Q*) Process

The ARMA(*p,q*) model is represented as:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \ldots + \phi_p y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \ldots - \theta_q \epsilon_{t-q}$$

Where $\{y_t\}$ is the data of daily COVID-19 deaths in India and $\phi_1, \phi_2, .., \theta_1, \theta_2$..are coefficients[251]

k. ARIMA(*p,d,q*) Process

The ARIMA (*p,d,q*) is represented as

$$y'_t = c + \phi_1 y'_{t-1} + \phi_2 y'_{t-2} + \phi_3 y'_{t-3} + \ldots + \phi_p y'_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \ldots - \theta_q \epsilon_{t-q}$$

Where $\{y_t\}$ is the data of daily COVID-19 deaths in India and $\phi_1, \phi_2, .., \theta_1, \theta_2$..are coefficients and y'_t is the differenced series, the predictors on the right-hand side include both lagged values ϵ_t and lagged errors.[252]

When neither the autocorrelations nor the partial autocorrelations are cut off, a mixed model is suggested. In an ARIMA(p,d,q) model, the autocorrelation function will be a mixture of exponential decay and damped sine waves after the first q-p lags. The partial autocorrelation function has the same pattern after p-q lags. By studying the first few correlations of each ACF and PACF plot, we obtain reasonable guesses for p and q. Later developments have led to other model selection tools such as the Akaike Information Criterion. The AIC is defined as,

$$AIC = -2(\text{log-likelihood}) + 2K$$

where K is the number of estimated model parameters.

l. Model Estimation

The ARIMA model is fitted to the resulting time series by using an automated process. Automated means that it should do all steps automatically and does not need any judgment or modification by humans. Easy application is satisfied when it only asks for the time series data as the input and gives its future forecast as output. Also, it should have accurate results and forecast in a reasonable time. The Box–Jenkins framework is extended to create an automated and integrated framework that supports dealing with complex situations, adjusts automatically with the stationarity of data, and provides reliable information to decision-makers in a reasonable time. The model estimated by the auto ARIMA method gives the lowest AIC value. The lowest value of AIC indicates a more accurate model with the best fit.[253]

m. Diagnosis Checking

After model estimation, diagnostic checking of the model is conducted which involves residual analysis. The model adequacy is primarily related to the assumption that residuals should be independent and identically distributed. Also, the residuals should be uncorrelated. They were examined by using the autocorrelation function (ACF), Partial autocorrelation function (PACF), ADF test, and Ljung- Box test.[254]

n. Residual Analysis

The residual in a time series model is the difference between the observations and the corresponding fitted values. The residuals generated by the model for the corresponding values of $wt = \nabla dxt$ are denoted by εt , t = 1,2, ...,n. Here, it is assumed that the unobserved residuals are normally distributed with zero mean and common variance, that is, $\varepsilon t \backsim N(0, \sigma 2)$. The first of its significance is that it provides a

diagnostic procedure for checking whether the initial specification of the model is correct. The expectation is that the residuals should resemble a white noise process which by assumption, are un-autocorrelated. If they are autocorrelated, new specifications are given for p, d and q and Ljung-Box-Pierce (Q) Test, Akaike Information Criteria (AIC), Standard Error (SE) of the time series model and Mean Absolute Percentage Error (MAPE) are considered and applied to determine the specification that best model the series under study.[255]

o. Ljung Box Test

Ljung Box test is a diagnostic tool to test the lack of fit of time series models. The test examines the autocorrelation of residuals and if the autocorrelation is very small it is clear that the model adequately fits the data. It considered the large sample properties of all the residual autocorrelated coefficients for any ARIMA process.[256] Here,

H_0: The model does not exhibit a lack of fit

H_1: The model exhibits a lack of fit

The test statistic is defined as follows;

$$Q = n\,(n + 2) \sum_{k=1}^{m} \frac{\hat{r}^2}{n-k}$$

Where n is the length of the time series, $\hat{r}$ is the estimated autocorrelation at lag k and m is the no of lags being tested. The test rejects the null hypothesis if $Q > \chi^2_{(1-\alpha,h)}$ where $\chi^2_{(1-\alpha,h)}$ is the Chi-square value with $h = m - p - q$ degrees of freedom and α level of significance (where p and q are the parameters of ARMA model fit to the data).

p. Forecasting

After the model identification of the ARIMA process, model estimation and diagnostic checking, the fitted model is used for forecasting the time series data. When a satisfactory model is found, forecasts are readily computed. Forecasts of future deaths are done using the death observation and the fitted value of residuals[257]

b. Compartmental models

Compartmental models were adapted to reflect the different waves of a pandemic like COVID-19, with each wave representing a distinct phase of transmission intensity and patterns. In this context, the compartments are divided according to the specific characteristics of each wave. For instance, during the first wave, the model focuses on compartments representing imported cases and initial local transmission. The second wave, often more intense and community-driven, would have compartments that reflect increased local transmission and possibly new virus variants.[258]

In each wave, the dynamics within the compartments—such as the number of susceptible, exposed, infected, and recovered individuals—change according to the specific conditions of that wave. Factors like public health interventions, population behaviour, and virus characteristics influence these dynamics. By analysing these compartmental transitions across different waves, public health officials and researchers gained insights into the evolving nature of the pandemic and tailored their response strategies accordingly.

The state-wise trends of COVID-19 Active Cases have been studied in South Indian states.

$$Active\ cases = Total\ Cases - (Cured + Death)$$

The case Fatality Rate of COVID-19 given as

$$CFR = \frac{Deaths}{Cured + Deaths} \times 100$$

The data on the number of confirmed cases, cured, and deaths available in the public domain from January 30, 2020, to May 19, 2021. The data from the day when deaths due to COVID-19 were more than one was considered for the construction of the model. Since the active cases for some states/ union territories are not linearly related, transformations are used to make it easier to model and analyse data.

A quadratic regression model is a polynomial of degree 2 with all the terms present.

$$EY_t = \beta_0 + \beta_1 t + \beta_2 t^2$$

This equation has the linear effect parameter β1 and quadratic effect parameter β2, respectively as well as the constant parameter β0.

This section of the study proposes a compartmental model for the dynamics of COVID-19 in south India. First of all, the data was normalised to the total population of individuals, and the compartments were created (Phase One, Resting Phase and Phase Two) as per trends in Active Cases. The entire period of COVID-19 in south Indian states has been divided into three groups. More specifically, the compartment collects all the individuals that are affected by the virus. Phase one is defined by the trend in active cases from the initial period, which has at least one extreme peak to the lowest value of the active cases. The resting phase is considered a linear, less drastic

period with an approximately constant magnitude in active cases. Phase two initiates from the sudden increase in the active cases from the restating phase.

c. Spatial Epidemiology and Forecasting

Crude Mortality Rate (CMR) and Case-fatality Rate (CFR) are significant indicators in epidemiology and public health that help to understand the impact of certain diseases, disasters, or public health scenarios. The analysis of these rates over time and across countries is crucial for policy-making and international health comparisons.[259]

- **Time series analysis and correlation**: A time series is a sequence of data points, typically consisting of successive measurements made over a time interval. Time series analysis comprises methods for analyzing time series data to extract meaningful patterns, statistics, and other characteristics of the data. In the context of CMR and CFR, it helps in understanding the trend, seasonality, and cyclical patterns. Correlation is a statistical measure that describes the degree to which two variables move in relation to each other. By studying the correlation between CMR or CFR and other variables (like economic indicators, healthcare system efficiency, etc.), the important factors affecting mortality rates were deduced.[260]
- **Autoregressive (AR) models**: An autoregressive model is a representation of a type of random process; as such, it describes certain time-varying processes in nature, including time series that show serial autocorrelation. In the context of CMR and CFR, AR models were useful for predicting future rates based on past values. For example, an AR(1) model would predict the CMR for a given year based on the rate from the previous year.[261]

- **Forecasting methods**: Forecasting is vital to make informed decisions based on future predictions of CMR and CFR. While AR models provide one method, there are other techniques like Moving Averages, and Exponential Smoothing. The choice of method often depends on the data available, the forecasting horizon, and the required precision.[262]
- **Comparison between countries**: Comparing CMR and CFR between countries provides insights into the effectiveness of healthcare systems, disease control measures, and overall health of populations. Differences in rates arose from variations in healthcare infrastructure, policies, public health interventions, and sociodemographic factors. Comparisons between countries further highlight areas of success and potential improvement in global health scenarios.[263]

2.3.4. Making Decisions

2.3.3.1 Pollution Controllability

Air pollution, characterized by high concentrations of particulate matter (PM) and nitrogen dioxide (NO2), is a pervasive environmental issue with known adverse effects on respiratory health. Numerous studies have established a link between prolonged exposure to air pollutants and an increased risk of respiratory infections and diseases. The respiratory system's compromised defines mechanisms under the influence of air pollution make individuals more susceptible to respiratory illnesses, including viral infections. In light of the COVID-19 pandemic, understanding the potential association between air pollution and the incidence of COVID-19 is of paramount importance, as it offers insights into environmental factors influencing the spread and severity of the virus.[264]

This section of the study aims to assess the air quality during the pre-lockdown and lockdown periods in India, focusing on the concentrations of particulate matter (PM) and nitrogen dioxide (NO2). Firstly, this section provides a comprehensive understanding of how environmental conditions, specifically air pollution levels, have fluctuated in response to the lockdown imposed due to the COVID-19 pandemic. By comparing pre-lockdown and lockdown air quality data, it was identified the extent of changes and variations in PM and NO2 concentrations.[265, 266]

The data collected from the Central Pollution Control Board included variable data along with the value of chemical pollutants PM_{10} (Particulate Matter less than 10 µm in diameter), $PM_{2.5}$ (Particulate Matter less than 2.5 µm in diameter), Nitrogen Dioxide (NO_2), and Sulphur Dioxide (SO_2) for a duration of 2 months pre and during the lockdown.

a. Air Quality Indices

The monitoring of ambient air quality was carried out using different air quality indices (AQIs). An AQI was defined as an overall value that converts weighted concentrations of individual air pollution-related parameters (SO2, Particulate Matters, CO, NO2, etc.) into a single number or set of numbers. This toxicity of the air is evaluated by classifying the AQIs according to the predefined criterion. In the present study, we consider four methods of calculating Air Quality Indices for two time periods which are calculated for each of the selected cities separately.[267]

Method 1 (AQI_{mean})**:** Ambient AQI is calculated by taking the arithmetic mean of the sum of the ratios of the four pollutants (PM_{10}, $PM_{2.5}$, CO_2 and SO_2) to their standard values and is multiplied by 100. These values are compared using a quality scale provided by CPCB. Here AQI for an individual pollutant, Q is given by Q = ($\frac{C}{Cs}$) * 100 where C is the observed value of the air quality parameters pollutant (PM_{10}, PM2.5, NO2 and SO2) and Cs is the CPCB standard for the given area.

Method 2 (AQI_{gm})**:** Ambient AQI is calculated by taking the geometric mean of the AQIs provided by individual components (as in Method 1) and is multiplied by 100. This measure is also compared with the quality scale provided by CPCB.

Method 3 (ORNAQI)**:** The Oak Ridge National Air Quality Index (ORNAQI) developed by the Oak Ridge National Laboratory (ORNL), USA is calculated using the mathematical formula AQI = [39.02 $\sum$ $\frac{Xi}{Xs}$] $^{0.967}$ where X_i is the value of individual air quality parameters and X_s is the prescribed standard value for that parameter. Air Quality Indexes measured by this method are then compared with relative ORAQI values.

Method 4 (AQI_{WeiAv})**:** This AQI is obtained by combining qualitative measures with the qualitative concept of the environment. The individual AQI is calculated as Q = $\frac{W*C}{Cs}$ where W is the weightage of the pollutant, C is the observed value of the pollutant and Cs is the CPCB standard for the given area. Here all the individuals are given with equal weight (W=1) and the total index is obtained as AQI = $\sqrt{(1/N)\sum_{i=1}^{N} \ Q\hat{}2}$ where N is the number of air quality variables.

b. Time Series Analysis

A time series is a record of values of any fluctuating quantity measured at different points in time. A Box-Jenkins approach to Time Series Analysis is used for the forecast of future data points of air pollutants 4 weeks posterior to the date. A Box-Jenkins model involves identifying an appropriate ARIMA process, fitting it to the date, and then using the fitted model for forecasting .[268]

c. Data Preparation

A preliminary Box-Jenkins analysis with a plot of the initial data should be run as the starting point in determining an appropriate model. The first, and most important, step in any time-series analysis is to plot the observations against time. This graph, called a *time plot*, will show important features of the series such as trends, seasonality, outliers and discontinuities. The plot is vital, both to describe the data and to help in formulating a sensible model. The input data should be adjusted to form a stationary series. Broadly speaking a time series is said to be stationary if there is no systematic change in mean (no trend), if there is no systematic change invariance and if strictly periodic variations have been removed. The stationarity of data is evaluated using the Augmented Dicky Fuller test. The apparent trends were adjusted by having a model apply a technique of regular differencing, a process of computing the difference between every two successive values, computing differenced series which have overall trend behaviour removed. If a single differencing does not achieve stationarity, it the process will be repeated.[269]

d. Autocorrelation and the Correlogram

An important guide to the properties of a time series is provided by a series of quantities called the sample autocorrelation coefficients. They measure the correlation, if any, between observations at different distances apart and provide useful descriptive information. We have seen that they are also an important tool in model building, and often provide valuable clues to a suitable probability model for a given set of data.[270]

Given $X(t_1)$, $X(t_2)$…, $X(t_n)$ be a time series observations for a set of times t_1, t_2…,t_n, then autocorrelation of order *h* is defined as;

$$\rho(h) = \frac{Cov(Xt, Xt-h)}{\sqrt{Var(Xt)Var(Xt-h)}}$$

This formula giving the autocorrelation coefficients is known as Autocorrelation Function (ACF) and a plot of $\rho(h)$ against h is known as auto correlogram and is often a good guide to the properties of time series.

e. Augmented Dicky Fuller Test

The Dickey-Fuller Test is one of the best-known and most commonly used unit root tests. It is based on the model of the first-order autoregressive process with an autoregressive parameter ϕ_1. The null hypothesis is, H_0: $\phi_1 \neq 1$ i.e. the process contains a unit root and therefore it is non-stationary and the alternative hypothesis is, H_1: $\phi_1 = 1$ i.e. the process does not contain a unit root and is stationary. The following equation is used to calculate the test statistic of the ADF test.[271]

$$\Delta y_t = (\phi_1 - 1)\, y_{t-1} + \sum_{i=1}^{p-1} \phi i(\Delta yt - i) + €_t$$

The level of significance (LOF) is taken as 0.05. If the p-value exceeds LOF, we accept the null hypothesis and if the p-value lies within the LOF, we reject the null hypothesis that the data is not stationary.

f. KPSS Test

The Kwiatkowski- Philips- Schmidt- Shin (KPSS) test is a type of unit root test that determines the stationarity of time series data around a deterministic trend. The test is built by the assumption of breaking the time series data into three parts; deterministic trend, random walk and stationary random error .[272] The test is based on the model;

$$y_t = d_t + r_t + €_t \text{ and } r_t = r_{t-1} + u_t$$

where $d_t = \sum_{i=0}^{p} \beta i t^i$, for p = 0,1 contains the deterministic part of the model, $€_t$ are IID $N(0,\sigma_€^2)$, r_t is a random walk with variance σ_u^2 and u_t are IID $N(0, \sigma_u^2)$.

KPSS test is based on Large Multiplier (LM) test with the hypothesis that the random walk has a zero variance i.e. H_0: $\sigma_u^2 = 0$ against the alternative that H_0: $\sigma_u^2 > 0$.

$$LM = \sum_{t=1}^{T} \frac{St\text{^}2}{\sigma€\text{^}2}$$

Where $s_t = \sum_{t=1}^{T} €t$, t = 1, 2, …, T and $\sigma_€^2$ is the estimated variance.

The $p < 0.05$ was considered as the significance level. Thus, the hypothesis H_0 of stationarity will be rejected if the observed value is less than 0.05.

g. Identification of Basic Model

With a stationary series in place, a basic model was identified. Three basic models exist, Autoregressive (AR), Moving Average (MA), and a combined in addition to the previously specified regular differencing combine to provide the available tools. When regular differencing is applied together with AR and MA, they are referred to as ARIMA, with the "I" indicating "Integrated" and referencing the differencing procedure.[273]

h. Autoregressive Processes

Suppose that $\{Z_t\}$ is a purely random process with mean zero and variance σ_z^2. Then a process $\{X_t\}$ is said to be an autoregressive process of order *p* (abbreviated to an AR(*p*) process) if

$$Xt = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{p-1} + Z_t$$

Where Z is a constant and α_i, I = 1, 2,…*p* are the model parameters.

This is rather like a multiple regression model, but X_t is regressed on past values of X_t rather than on separate predictor variables.[274]

i. Moving Average Processes

Suppose that $\{Z_t\}$ is a purely random process with mean zero and variance σ_z^2. Then a process {Xt} is said to be a moving average process of order *q* (abbreviated to an MA(*q*) process) if

$$X_t = \beta_0 Z_t + \beta_1 Z_{t-1} - \beta_2 Z_{t-2} - \ldots - \beta_j Z_{t-q}$$

Where β_j, j = 1, 2,…*q* are the model parameters. The Zs are usually scaled so that β_0=1.[275]

j. Mixed ARMA Models

A useful class of models for time series is formed by combining MA and AR processes. A mixed autoregressive/moving-average process containing *p* AR terms and *q* MA terms is said to be an ARMA process of order (*p*, *q*). It is given by

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + Z_t - \beta_1 Z_{t-1} - \beta_2 Z_{t-2} - \ldots - \beta_j Z_{t-q}$$

Where α_i, I = 1, 2,…*p* and β_j, j = 1, 2,…*q* are the AR and MA model parameters respectively.

The importance of ARMA processes lies in the fact that a stationary time series is often adequately modelled by an ARMA model involving fewer parameters than a pure MA or AR process by itself. This is what is often called the Principle of Parsimony. This says that we want to find a model with as few parameters as possible, but which gives an adequate representation of the data at hand.[276]

k. Integrated ARMA (or ARIMA) Models

In practice most time series are non-stationary. To fit a stationary model, it is necessary to remove non-stationary sources of variation. If the observed time series is non-stationary in the mean, then the series will be differentiated.

If X_t is replaced by $\Delta_d X_t$ in the equation, then we have a model capable of describing certain types of non-stationary series. Such a model is called an 'integrated' model because the stationary model that is fitted to the differenced data has to be summed or 'integrated' to provide a model for the original non-stationary data.[277]

$$\Delta_d X_t = W_t = (1 - B)^d X_t$$

The general autoregressive integrated moving average (ARIMA) process is of the form

$$Wt = \alpha_1 W_{t-1} + \ldots + \alpha_p W_{t-p} + Z_t + \ldots + \beta_q Z_{t-q}$$

Rewritten (a) as equation (b):

$$Y_t = (1 - \alpha_1) Y_{t-1} + (\alpha_2 - \alpha_1) Y_{t-2} + (\alpha_3 - \alpha_2) Y_{t-3} + \ldots + (\alpha_p - \alpha_{p-1}) Y_{t-p} - \alpha_p Y_{t-p-1} + Z_t - \beta_1 Z_{t-1} - \beta_2 Z_{t-2} - \ldots - \beta_q Z_{t-q}$$

Where α_i, $i = 1, 2, \ldots p$ and β_j, $j = 1, 2, \ldots q$ are the AR and MA model parameters respectively

Using the backward shift operator B,

$$\phi(B) (1 - B)^d X_t = \theta(B) Z_t$$

Where $\phi(B)$, $\theta(B)$ are polynomials of order p, q, respectively.

In practice, first differencing is often found to be adequate to make a series stationary, and so the value of d is often taken to be one.

Model selection in Box-Jenkins uses various graphs based on transformed and differenced data to try to identify potential ARIMA processes that provide a good fit to the data. Later developments have led to other model selection tools such as Akaike Information Criterion.

l. Akaike Information Criterion

The Akaike Information Criterion allows testing how well the model fits the dataset without overfitting it. The accuracy and performance of the model are measured using AIC and is given (approximately) by:

$$AIC = -2 \ln(\text{max. likelihood}) + 2p$$

Where p denotes the number of independent parameters estimated in the model. Thus, the AIC essentially chooses a model with the best fit, as measured by the likelihood function, subject to a penalty term that increases with some parameters fitted in the model. This should prevent overfitting.[278]

m. Estimation of Parameters

The major diagnostic tool used for fitting a suitable model to an observed time series is the sample Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). Inference based on this function is often called analysis in the time domain.[279]

n. Modelling MA Processes Using ACF

The correlogram helps try to identify a suitable class of models for a given time series, and, in particular, for selecting the most appropriate type of autoregressive integrated moving average (ARIMA) model. A correlogram does not come down to zero reasonably quickly, indicates non-stationarity and so the series needs to be differenced. For stationary series, the correlogram is compared with the theoretical ACFs of different ARMA processes to choose the one that seems to be the 'best' representation. The ACF of an MA(q) process is easy to recognize as it 'cuts off' at lag q.[280]

o. Modelling AR Processes Using PACF

It is often difficult to assess the order of an AR process from the sample ACF alone. For a first-order process, the theoretical ACF decreases exponentially and the

sample function should have a similar shape. However, for higher-order processes, the ACF, a mixture of damped exponential or sinusoidal functions which is difficult to identify. To identify the order of an AR process we use the concept of Partial Autocorrelation Function (PACF). When fitting an AR(p) model, the last coefficient α_p will be denoted by π_p and measures the excess correlation at lag p which is not accounted for by an AR($p-1$) model. It is called the p^{th} partial autocorrelation coefficient and, when plotted against p, gives the PACF. The first partial autocorrelation coefficient π_1 is simply equal to $\rho(1)$, and this is equal to α_1 for an AR(1) process.[281]

p. Model Adequacy Checking

The model checking involves testing the assumptions of the model to identify the areas where the model is inadequate.

I. Residual Analysis

To estimate the model adequacy, residual analysis is also performed. The residuals in the time series model are what is left over after fitting the model.[282] The residual generated by the model is equal to the difference between the observations and the corresponding fitted value.

$$e_t = y_t - \hat{y}_t$$

Where y_t is the observed value of model parameters and $\hat{y}_t$ is the fitted value.

If we have a 'good' model, then we expect the residuals to be 'random' and 'close to zero', and model validation usually consists of plotting residuals using time plots and correlograms. For an adequate model, the residuals should be Independent and Identically Distributed (IID) and are uncorrelated. To test the correlation, we use

Auto Correlation Function (ACF) plot and the normality of the residuals is tested using the Shapiro-Wilk Test of Normality.

II. Shapiro Wilk Test

Shapiro Wilk test is a general normality test designed to test whether the selected sample belongs to a normal population or not. The null hypothesis is that the selected sample x_1, x_2, …x_n belongs to a normal population. The W test statistic is given by;

$$W = \frac{(\sum_{i=1}^{n} a_i x_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Where n is the sample size, x_i is the ordered sample values and a_i are the constants generated from means, variances and covariance from the order statistics from a normal population and μ is the sample mean. The level of significance is taken as 0.05. If p-value < 0.05, the null hypothesis is rejected and was concluded that the test data is not normally distributed.[283]

III. Ljung Box Test

Ljung Box test is a diagnostic tool to test the lack of fit of time series models. The test examines the autocorrelation of residuals and if the autocorrelation is very small it is clear that the model adequately fits the data. Here the null hypothesis, H_0: the model does not exhibit a lack of fit is tested against the alternative hypothesis, H_1: the model exhibits a lack of fit and the test statistic is defined as follows;

$$Q = n\,(n + 2) \sum_{k=1}^{m} \frac{\hat{\imath}\text{^}2}{n-k}$$

Where n is the length of the time series, r̂ is the estimated autocorrelation at lag k and m is the no of lags being tested. The test rejects the null hypothesis if Q > $\chi^2_{(1-\alpha,h)}$ where $\chi^2_{(1-\alpha,h)}$ is the Chi-square value with h = m – *p* – *q* degrees of freedom and α level of significance (where *p* and *q* are the parameters of ARMA model fit to the data).[284]

I. Forecasting

Forecasting is what the whole procedure is designed to accomplish. Once the model has been selected, estimated and checked, it is usually a straightforward task to forecast. When a satisfactory model is established, it becomes easy to generate forecasts. Given data up to time n, these forecasts will involve the observations and the fitted residuals (i.e. the one-step-ahead forecast errors) up to and including time n.[285]

i. Forecast Error

Forecast error is the difference between the observed value and its forecast. It should be noted that forecast errors are different from residuals since the forecast errors are based on a test set and it involve multi step forecasts. The forecast errors were summarized to measure forecast accuracy in different ways.[286] The forecast error, e_t is given by

$$e_t = y_t - \hat{y}_t$$

where y_t is the observation and $\hat{y}_t$ is the forecast of the query.

ii. Mean Absolute Error

Mean Absolute error (MAE) is the scale dependent accuracy measure used for comparing forecast methods applied to a single time series or to several time series with the same units. It calculates the absolute difference between the predicted value and actual value of each query. MAE identifies how big of an error we expect from the forecast of an average.[287]

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |et|$$

Where n is the total number of queries and e_t is the forecast error.

iii. Mean Absolute Percentage Error

Mean Absolute Percentage Error (MAPE) are percentage error which are unit free and used to compare the forecast experiences. It measures the accuracy of the forecasting system as the average of absolute percent error. It reports higher errors when the forecast is greater than the actual value and lower rates if the forecast is less than actual value.[288]

$$MAPE = 100 * \frac{1}{n}\sum_{i=1}^{n} \left|\frac{et}{yt}\right|$$

Where n is the total number of queries and e_t is the forecast error.

iv. Paired t-Test

Paired *t*-test was used to ascertain the significance of differences between mean values of individual pollutants during pre and during lockdown for the selected cities. When the samples X and Y are independently and normally distributed with common variance and respective sample sizes n and m and means μ_1 and μ_2, hypothesis about their means was tested as follows. Here the null hypothesis, the true

mean difference between the paired samples is zero, is tested against the alternative hypothesis that the true mean difference of the paired samples is not equal to zero.[289] Then a test was found based on the statistic:

$$t = \frac{\mu_1 - \mu_2}{S_P\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

$$\text{Where } S_p = \sqrt{\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}}$$

Now, T has a t distribution with r = n + m – 1 degree of freedom. The level $p < 0.05$ was considered as the significance level. Thus, the hypothesis H_0 will be rejected in favour of H_1 if the observed value of t is less than 0.05.

v. Pearson Coefficient of Correlation

The Pearson Correlation Coefficient is used to identify the correlation between the individual air pollutants in the selected geographic areas.[290]

Karl Pearson (1867-1936), the famous British statistician has proposed a coefficient measure that measures the degree of relationship between two variables. We assume that there is a linear relationship between two random variables x and y.

Then Pearson correlation coefficient,

$$r_{xy} = \frac{p_{xy}}{\sigma_X \sigma_y}$$

Where p_{xy} is the product-moment or covariance between x and y, σ_x^2 and σ_y^2 are the variances of x and y respectively. This was interpreted as follows;

If $r_{xy} > 0$; there exists a positive correlation between the two variables.

If $r_{xy} = 0$; there exists no correlation between the two variables.

If $r_{xy} < 0$; there exists a negative correlation between the two variables.

2.3.3.2 Optimal allocation of limited resources

The study of optimal allocation of limited resources in the context of the COVID-19 pandemic, where resources such as medical supplies, healthcare personnel, and vaccines are often scarce compared to the needs of the population. During the pandemic, healthcare systems around the world have faced unprecedented challenges in managing resources effectively. Decisions on how to allocate limited ventilators, intensive care unit (ICU) beds, personal protective equipment (PPE), and later, vaccines, have had profound implications on patient outcomes and the overall control of the virus spread. The study of optimal resource allocation in this scenario helps in developing strategies that ensure the most effective use of resources, prioritizing areas and populations most in need, and adapting to the changing dynamics of the pandemic. This not only improves patient care and survival rates but also helps in mitigating the broader public health impacts of the pandemic.[291]

Moreover, the COVID-19 pandemic has underscored the need for equitable resource distribution, both within and between countries. The challenge of allocating limited resources extends beyond hospitals to encompass the global distribution of vaccines and medical aid. This aspect of the study focuses on ethical considerations, prioritizing high-risk groups, and addressing disparities in healthcare access. It also involves strategizing the distribution logistics to ensure timely and fair access to resources across different regions. By optimizing resource allocation, the response to

the pandemic was more effective, reducing the overall burden of the disease and paving the way for a more coordinated and equitable global health response.[292]

i. Compare the Availability and Demand of Beds, ICU, and Ventilators

Comparing the availability and demand of beds, ICU units, and ventilators is essential in healthcare resource management, particularly during crises like the COVID-19 pandemic. General hospital beds, which are fundamental for patient care, were in high demand as COVID-19 cases surged, often leading to shortages and the need for rapid expansion of capacity. ICU units, equipped for severe cases requiring intensive care, also faced significant strain. The demand for these specialized units often exceeded availability, necessitating the conversion of other hospital areas into temporary ICU spaces. Ventilators, crucial for patients with severe respiratory issues, became a critical and scarce resource. The imbalance between the limited supply of ventilators and the high number of patients needing respiratory support posed a major challenge, leading to difficult triage decisions and an urgent call for increased production and distribution. This comparison highlights the stress on healthcare systems during the pandemic and underscores the importance of preparedness and flexible resource allocation.[293]

ii. Predicting Time to Overload the Available Facilities

Predicting the time to overload available healthcare facilities, such as hospital beds, ICU units, and ventilators, is crucial for effective pandemic management. This involves analyzing current resource usage trends, infection rates, and healthcare capacities to forecast when these facilities reach their limits. The prediction process typically relies on epidemiological models that factor in the rate of COVID-19

transmission, the proportion of cases requiring hospitalization, ICU care, or ventilators, and the duration of hospital stays. By inputting current data on infection rates and healthcare utilization, these models were estimated when healthcare facilities are likely to be overwhelmed. For instance, if the number of new daily cases is increasing rapidly and a significant percentage of these cases require hospitalization, models projected the point in time when hospital beds or ICU units will reach full capacity.[294]

Additionally, these predictions should consider variables such as changes in public health policies, vaccination rates, and the emergence of new virus variants, which significantly alter transmission dynamics. Hospitals and health systems also use these predictions to implement contingency plans, such as increasing bed capacity, setting up temporary medical facilities, or reallocating medical staff and equipment. Accurate predictions are vital for ensuring that healthcare systems prepare for and mitigate the impacts of facility overloads, ultimately saving lives and maintaining the quality of healthcare delivery during pandemic peaks.[295]

iii. Dynamics of Virus Spread and Impact on Healthcare Systems

The spread of a virus like COVID-19 is intricately linked to its reproductive number, commonly referred to as R_0 (basic reproduction number). This number represents the average number of new infections generated by an infected individual in a fully susceptible population. An R_0 value greater than 1 indicates that the virus is spreading rapidly. Throughout the COVID-19 pandemic, the R_0 has varied, influenced by factors such as population density, public health measures like social distancing and mask mandates, and the emergence of more transmissible variants. For example, the original strain of COVID-19 had a lower R_0 compared to later variants

like Delta and Omicron, which were more infectious and altered the dynamics of virus transmission. Monitoring and understanding the R_0 is crucial for predicting the spread of the virus and implementing timely interventions to curb its transmission.[296]

The saturation of healthcare systems during the pandemic has been a direct consequence of this rapid spread of the virus. As the R_0 rises and more individuals become infected, the demand for healthcare resources such as hospital beds, ICU units, and ventilators increases dramatically. Many healthcare systems around the world have faced overwhelming challenges during peaks of COVID-19 cases, with hospitals operating at or beyond their capacity. This saturation not only impacts the care of COVID-19 patients but also strains the overall healthcare provision, affecting treatments for other medical conditions. Managing the spread of the virus to keep the R_0 at a manageable level is vital to prevent healthcare system overload. This involves a combination of public health policies, vaccination campaigns, and community engagement to reduce transmission and ensure that healthcare systems provide adequate care to all patients.[297]

2.3.3.3 Impact Assessment Tool

a. Social Impact

The secondary data from a cross-sectional study which was designed to better understand the level of psychological impact, anxiety, depression and stress among the residents of Kerala during the outbreak of COVID-19 was used for assessing the social impact. Information on socio demographic data, knowledge, and concern about COVID-19 and precautionary measures against COVID-19 were documented. The socio-demographic data includes age, gender, education, residential location,

employment status, marital status, annual income and household size. Data was collected in the month of May 2020. The data was collected from different districts of Kerala.[298]

Mental health status was assessed by Depression Anxiety Stress Scale and psychological impact (Post Traumatic Stress Disorder symptoms) was assessed by the Impact of Event Scale-Revised. The Depression Anxiety Stress Scale is a set of three self-report scales which include a total of 21 items (DASS-21). Each of the three DASS-21 scales is divided into subscales containing 7 items each with similar content. The depression scale assesses dysphoria, hopelessness, devaluation of life, self-deprecation, lack of interest/ involvement, anhedonia and inertia. The anxiety scale assesses autonomic arousal, skeletal muscle effects, situational anxiety, and subjective experience of anxious affect. The stress scale is sensitive to levels of chronic non- specific arousal. It assesses difficulty relaxing, nervous arousal, and being easily upset/agitated, irritable/over-reactive and impatient. Scores for depression anxiety and stress are calculated by summing the scores for the relevant items.[299]

I. Depression Anxiety Stress Scale (DASS)

The Depression Anxiety Stress Scale, commonly known as DASS, is a clinical assessment tool designed to measure the three related negative emotional states of depression, anxiety, and stress. Its scoring is divided into specific ranges, which classify the severity of each emotional state. For depression, scores between 0-9 are considered normal, 10-12 are classified as mild, 13-20 as moderate, and 21-42 indicate severe depression. For anxiety, the corresponding ranges are 0-6 (normal), 7-9 (mild), 10-14 (moderate), and 15-42 (severe). Lastly, for stress, scores of 0-10 are

deemed normal, 11-18 as mild, 19-26 as moderate, and 27-42 as severe. The specific score ranges serve as a guide for clinicians to evaluate the intensity of each emotional state in an individual and to determine appropriate therapeutic interventions.[300]

The Impact of Event Scale-Revised is a self-administered questionnaire consisting of 22 questions with three subscales intrusion, avoidance and hyperarousal. These subscales measure different dimensions of the stress scale. The response set of intrusion measures (intrusive thoughts, nightmares, intrusive feelings and imagery) and avoidance measures (numbing of responsiveness, avoidance of feelings, situations and ideas) and hyper arousal measures (anger, irritability, hyper-vigilance, difficulty in concentrating, and heightened startle). While there is no cut-off score, scores more than 24 are of concern.[301]

II. IES-R Scale Explanation

The IES-R (Impact of Event Scale – Revised) is a widely used instrument for assessing subjective distress caused by traumatic events. It provides a measure of the severity of a person's response to a specific traumatic event. The scale is divided into different severity levels based on the total score a person receives. Scores ranging from 0 to 23 are considered "Normal", indicating minimal distress related to the event. Scores from 24 to 32 fall under the "Mild" category, suggesting a mild level of distress. When scores lie between 33 and 36, they're categorized as "Moderate", implying a moderate level of distress. Finally, a score of 37 and above is classified as "Severe", indicating a significant amount of distress associated with the traumatic event. Thus, the IES-R scale serves as a useful tool for professionals to gauge the psychological impact of trauma and guide potential interventions or treatments.[302]

The scores obtained from the above scales are used to assess the psychological impact and mental health of Keralite during lockdown due to the outbreak of COVID-19.

III. List of Variables

a. Demographic Variables

- **District:** District where the participants are residing.
- **Age:** Age of the participants in the study.
- **Gender:** Gender of the participants in the study. It is classified as male, female and transgender.
- **Marital Status:** The marital status of the participants is classified as single and married.
- **Household Size:** Number of family members.
- **Educational Qualification:** The educational qualification of the participants involved in the study and is classified as plus two or below, graduate, post graduate.
- **Occupational Status**: The occupational status of the participants involved in the study is classified as student, unemployed, healthcare professional, govt. employee, and private employee.
- **Monthly Family Income**: The monthly family income of the participants is classified as below 5000, 5000-20000, 20000-50000, 50000-75000 and 75000 and above.
- **Loss Expecting Per Month**: Individual loss of the participant during lockdown period per month.
- **Area of Residence**: The residing area where the participants currently live and are classified as urban and rural.

b. Other variables

1. Variables related to precautionary measures:

- Ways through which information about hygiene is gained (friends, mass media (T.V, newspaper, social media).
- Social media exposure (never, sometimes, often, very often).
- Hygiene practices (washing hands, wearing mask, social distancing, covering mouth when coughing and sneezing).

2. Additional health information:

- Chronic diseases: Life style diseases such as diabetes, blood pressure.
- Extra activities: other activities apart from their regular routine work (doing meditation or yoga, using social media, doing something I am good at, spending time with loved ones.

IV. Tests involved

i. Chi Square Test

The Chi-square statistic is a non-parametric tool designed to analyze group differences when the dependent variable is measured at a nominal level. The Chi-square is robust with respect to the distribution of the data. Specifically, it does not require equality of variances among the study groups or homoscedasticity in the data. It permits evaluation of both dichotomous independent variables, and of multiple group studies. The calculations needed to compute the Chi-square provide

considerable information about how each of the groups performed in the study.[303] The Chi-square is a significant statistic, and should be followed with a strength statistic. Advantages of the Chi-square include its robustness with respect to distribution of the data, its ease of computation, the detailed information that was derived from the test, its use in studies for which parametric assumptions cannot be met, and its flexibility in handling data from both two group and multiple group studies. Limitations include its sample size requirements, difficulty of interpretation when there are large numbers of categories (20 or more) in the independent or dependent variables.[304]

The Chi-square test is a non-parametric statistic, also called a distribution free test. Non-parametric tests should be used when any one of the following conditions pertains to the data:

The level of measurement of all the variables is nominal or ordinal.

The sample sizes of the study groups are unequal; for the chi square the groups are of equal size or unequal size whereas some parametric tests require groups of equal or approximately equal size.

The original data were measured at an interval or ratio level, but violate one of the following assumptions of a parametric test:

- The distribution of the data was seriously skewed and thus the researcher should use a distribution free statistic rather than a parametric statistic.
- The data violate the assumptions of equal variance or homoscedasticity.
- For any of a number of reasons the continuous data were collapsed into a small number of categories, and thus the data are no longer interval or ratio.[305]

ii. **Assumptions of the Chi-Square**

The Chi-square assumes that the data were obtained through random selection. The assumptions of the Chi-square include:

- The data in the cells should be frequencies, or counts of cases rather than percentages or some other transformation of the data.
- The levels (or categories) of the variables are mutually exclusive. That is, a particular subject fits into one and only one level of each of the variables.
- Each subject contributes data to one and only one cell in the Chi-square.
- The study groups should be independent. This means that a different test should be used if the two groups are related.
- There are two variables and both are measured as categories, usually at the nominal level. However, data may be ordinal data. Interval or ratio data that have been collapsed into ordinal categories may also be used.
- The value of the cell *expected* should be 5 or more in at least 80% of the cells, and no cell should have an expected of less than one.[306]

To test H_0: the attributes A & B are independent.

Test statistics is, $$\chi^2=\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

Where O_{ij} : Observed frequency in the $(i,j)^{th}$ cell

E_{ij} : Expected frequency in the $(i,j)^{cell}$

When H_0 is true, where f_{ij} is the expected frequency under H_0 in the $(i,j)^{th}$ cell and is given by

$$E_i = \frac{R_i * C_j}{N}$$

Where, R_i: Total of the i^{th} row

C_j: Total of the j^{th} column

iii. Shapiro Wilk Test

Shapiro Wilk test is a general normality test designed to test whether the selected sample belongs to a normal population or not. The null hypothesis is that the selected sample $x_1, x_2, \ldots x_n$ belongs to a normal population. The test statistic is given by,

$$W = \frac{(\sum_{i=1}^{n} a_i x_i)^2}{\sum_{i=1}^{n}(x_i - \mu)^2}$$

Where n is the sample size, x_i is the ordered sample values and a_i are the constants generated from means, variances and covariance from the order statistics from a normal population and μ is the sample mean. The level of significance is taken as 0.05. If p-value < 0.05, the null hypothesis is rejected and was concluded that the test data is not normally distributed.[307]

iv. Mann-Whitney U Test

The Mann-Whitney U test is used to compare differences between two independent groups when the dependent variable is either ordinal or continuous, but not normally distributed. It is often considered the nonparametric alternative to the independent t-test although this is not always the case.[308]

Mann-Whitney U test allows to draw different conclusions about the data depending on the assumptions that are made about the data's distribution. These

conclusions range from simply stating whether the two populations differ through to determining if there are differences in medians between groups. These different conclusions hinge on the shape of the distributions of the data. In order for a Mann-Whitney U test to give a valid result it requires some assumptions that the data should meet.[309]

Assumptions

- The dependent variable should be measured at the ordinal or continuous level.
- The independent variable should consist of two categorical, independent groups.
- The independence of observations, which means that there is no relationship between the observations in each group or between the groups themselves.

A Mann-Whitney U test was used when the two variables are not normally distributed.

We consider two independent random samples X = $(X_1, X_2, \ldots, X_{n1})$ and

Y = $(Y_1, Y_2, , \ldots, Y_{n2})$ from two populations with observed values $(x_1, x_2, \ldots, x_{n1})$ and $(y_1, y_2, \ldots, y_{n2})$ respectively. In this case, the null hypothesis H_0can be formulated as

$$H_0: P(X > Y) = P(Y > X) = \frac{1}{2}$$

The null hypothesis can be interpreted as the probability that a randomly drawn observation from the first population has a value *x* that is greater (or lower) than the value *y* of a randomly drawn subject from the second population is$\frac{1}{2}$. The alternative hypothesis H_1 is then

$$H_1: P(X > Y) \neq P(Y > X) .$$

This test is often performed as a two-sided test and, thus, the research hypothesis indicates that the populations are not equal as opposed to specifying directionality. A one-sided research hypothesis is used if interest lies in detecting a positive or negative shift in one population as compared to the other. The procedure for the test involves pooling the observations from the two samples into one combined sample, keeping track of which sample each observation comes from, and then ranking lowest to highest from 1 to n_1+n_2, respectively.[310]

Test Statistic for the Mann Whitney U Test

The test statistic for the Mann Whitney U Test is denoted U and is the smaller of U_1 and U_2, defined below.[311]

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

Where R_1 = sum of the ranks for group 1 and R^2 = sum of the ranks for group 2.

v. Kruskal-Wallis test

The Kruskal-Wallis one-way analysis of variance by ranks is an extremely useful test for deciding whether k independent samples are from different populations.[312]

Assumptions

The samples are independent random samples from their respective populations.

The measurement scale employed is at least ordinal.

The distributions of the values in the sampled populations are identical except for the possibility that one or more of the populations are composed of values that tend to be larger than those of the other populations.

If there are k samples, and varying number n_i is studied in each sample let n = total studied = $\sum_{i=1}^{k} n_i$. Let all the values from all the k samples be combined and ranked in a single series. The smallest value is replaced by rank 1, the next to smallest by rank 2, and the largest by rank n. n = the total number of independent observations in the k samples. When this has been done, the sum of the ranks in each sample is found. The Kruskal-Wallis test determines whether these sums of ranks are so disparate that they are not likely to have come from samples which were all drawn from the same population. It was shown that if the k samples actually are from the same population or from identical populations, the H is distributed as Chi-square with df = k-1, provided that the sizes of the various k samples are not too small.

Test statistic is given by

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1)$$

Where, k = number of samples

n_j = number of observations in the j^{th} sample

n = number of observations in all samples combined

R_j = sum of the ranks in the j^{th} sample

vi. Pearson correlation coefficient

Correlation is a statistical measure that indicates the extent to which two or more variables fluctuate together. A positive correlation indicates the extent to which

those variables increase or decrease in parallel; a negative correlation indicates the extent to which one variable increases as the other decreases.[313]

When the fluctuation of one variable reliably predicts a similar fluctuation in another variable, there's often a tendency to think that means that the change in one causes the change in the other. However, correlation does not imply causation. There are unknown factors that influence both variables similarly.

Correlation is a statistical technique that shows whether and how strongly pairs of variables are related. Although this correlation is fairly obvious your data contains unsuspected correlations. The exploration of potential correlations within the data is an important aspect of this research. While there are indications of various relationships, identifying the most significant correlations remains a key focus of the analysis. An intelligent correlation analysis leaded to a greater understanding of the data.[314]

Correlation is positive or direct when the value increases together, and

Correlation is negative when one value decreases as the other increases, and so called inverse or contrary correlation

If the points plotted were all on a straight line we would have perfect correlation, but it could be positive or negative

Correlation have a value:

1 is a perfect positive correlation

0 is no correlation (the values don't seem linked at all)

-1 is a perfect negative correlation

The value shows how good the correlation is (not how steep the line is), and if it is positive or negative. Usually, in statistics, there are three types of correlations: Pearson correlation, Kendall rank correlation and Spearman correlation.

Assumptions of Correlation

Employing correlation relies on some underlying assumptions. The variables are assumed to be independent, assume that they have been randomly selected from the population; the two variables are normal distribution; the association of data is homoscedastic (homogeneous), homoscedastic data have the same standard deviation in different groups where data are heteroscedastic have different standard deviations in different groups and assumes that the relationship between the two variables is linear. The correlation coefficient is not satisfactory and difficult to interpret the associations between the variables in case if data have outliers.

An inspection of a scatter plot gave an impression of whether two variables are related and the direction of their relationship. But it alone is not sufficient to determine whether there is an association between two variables. The relationship depicted in the scatter plot needs to be described qualitatively. Descriptive statistics that express the degree of relation between two variables are called correlation coefficients. A commonly employed correlation coefficient is Pearson correlation, Kendall rank correlation and Spearman correlation.

Correlation used to examine the presence of a linear relationship between two variables providing certain assumptions about the data are satisfied. The results of the analysis, however, need to be interpreted with care, particularly when looking for a causal relationship.

Bivariate correlation is a measure of the relationship between the two variables; it measures the strength and direction of their relationship, the strength ranging from absolute value 1 to 0. The stronger the relationship, the closer the value is to 1. Direction of the relationship was positive (direct) or negative (inverse or contrary); correlation generally describes the effect that two or more phenomena occur together and therefore they are linked. For example, the positive relationship of .71 represented a positive correlation between the statistics degrees and the science degrees. The student who has a high degree in statistics has also a high degree in science and vice versa.[315] The Pearson correlation coefficient is given by the following equation:

$$r = \frac{\sum_{i=1}^{n}(xi - \bar{x})(yi - \bar{y})}{\sqrt{\sum_{i=1}^{n}(xi - \bar{x})^2 \sum_{i=1}^{n}(yi - \bar{y})^2}}$$

Where $\bar{x}$ is the mean of variable x values, and $\bar{y}$ is the mean of variable y values

vii. Median regression

Quantile regression constitutes a family of statistical techniques used to estimate and draw inferences about conditional quantile functions. The quantile regression method is an extension of the traditional least square regression method. The difference is that the former accurately estimated the marginal effect of independent variables on the dependent variable in a specific condition quantile, which is superior to the marginal effect of the latter's average trend. Quantile regression offers a systematic strategy for examining how covariates influence the location, scale, and shape of the entire response distribution. It quantifies the heterogeneous effects of covariates through conditional quantiles of the outcome variable, and provides a comprehensive scan of the whole distribution of the outcome.

Quantile regression complements and improves the traditional mean regression models. In this situation of assumption violated, quantile regression quantifies the heterogeneous effects of covariates through conditional quantiles of the outcome variable, and provides a comprehensive scan of the whole distribution of the outcome.[316]

Quantile regression does not restrict attention to the conditional mean and therefore it permits to approximate of the whole conditional distribution of a response variable. Quantile regression is desired if conditional quantile functions are of interest. The development and dissemination of quantile regression started with the formulation of the quantile regression problem as a linear programming problem. Such formulation allows exploiting efficient methods and algorithms to solve a complex optimization problem, offering a way to explore the whole conditional distribution of a variable and not only its centre. The quantile regression problem typically exploits a variant of the well-known simplex algorithm for a moderate size problem. Quantile regression is an invaluable tool for facing heteroscedasticity, and provides a method for modelling the rates of change in the response variable at multiple points of the distribution when such rates of change are different. Empirical conditional quantile function provides a standard distribution-free approach to obtain prediction intervals for population quantiles based on sample regression quantiles. Such use of regression quantile estimates has the advantage of robustness against departure from the normality assumption of the error terms. Although in many practical applications the focus is on estimating a subset of quantile regression estimates, it is possible to obtain estimates over the whole interval (0, 1) through the so-called quantile process.[317]

The quantile regression was described by equation:

$$y_i = x_i' + \beta_q + e_i$$

Where βq is the vectors of unknown parameters associated with q [th] quantile

The p[th] conditional quantile given $x_i$is Q(p) ($y_i|x_i$) = $\beta_0^p + \beta_{01}^p x_i$. Thus, the conditional p[th] quantile is determined by the quantile-specific parameters, $\beta_0^p$and β_1^p and a specific value of the covariate $x_{i.}$

Any real-valued random variable X, characterized by its (right continuous) distribution function,

F(x) = P(X ≤ x), for any 0 <τ <1,

$F^{-1}(\tau)$ = inf{x: F(x)≥ τ is called the τ^{th} quantile of X.

The quantile function, associated with a probability distribution of a random variable, specifies the value of the random variable such that the probability of the variable being less than or equal to that value equals the given probability.

The conventional mean regression minimizes the sum of squares of residuals.

$$\text{i.e.; } \min \sum_i \left(Y_i - (\beta_0 + \beta_1)\right)^2$$

whereas, median regression minimizes the sum of absolute residuals.

$$\text{i.e.; } \min \sum_i |e_i|$$

In the quantile regression, the distance of points from a line is measured using a weighted sum of vertical distances (without squaring), where the weight is 1 – p for points below the fitted line and p for points above the line. Each choice for this proportion p, for example, p = .10, .25, .50, gives rise to a different fitted conditional-

quantile function. The mean of a distribution was viewed as the point that minimizes the average squared distance over the population, whereas a quantile q was viewed as the point that minimizes an average weighted distance, with weights depending on whether the point is above or below the value q.

Let A = (x_1,...,x_n) denote the matrix consisting of n observed vectors of the random vector X, and let y = (y_1,...,y_n) denote the n observed responses. The model for linear quantile regression is

$$Y = A^{'}\beta$$

where $\theta = (\theta_1, \theta_2. . . , \theta_p)'$ is the unknown p-dimensional vector of parameters

and $\varepsilon = (\varepsilon_1, \varepsilon_2 \ldots \varepsilon_n)$ is the n dimensional vector of unknown errors.

The τ^{th} regression quantile is a solution:

$$\text{Min } \{\sum_{i\in(i:y_i>x_i'\beta)} \tau(y_i - x_i'\beta) + \sum_{i\in(i:y_i<x_i'\beta)}(1-\tau)(y_i - x_i'\beta)\}$$

When τ =1/2 it becomes median regression.

Median regression was formulated as a linear programming problem and was solved efficiently with simplex algorithm.

The median regression line should pass through a pair of data points with half of the remaining data lying above the regression line and the other half falling below. That is, roughly half of the residuals are positive and half are negative.[318]

viii. Features of Quantile Regression[319]

- The entire conditional distribution of the dependent variable Y can be characterized through different values of τ
- Heteroscedasticity can be detected.
- If the data is heteroscedastic, median regression estimators can be more efficient than mean regression estimation
- The minimization problem can be solved efficiently by linear programming methods, making estimation easy.
- Quantile functions are also equivariant to monotone transformations.
- Quantiles are robust in regard to outliers

CHAPTER 3: RESULTS

3.1 Monitoring

1.1.1 SWOT analysis of data-driven methods in COVID-19 pandemic

3.1.1.1 Strengths

The strengths of data-driven methods in tackling the COVID-19 pandemic were cohesively outlined in several key areas:

The experience gained from past epidemics such as SARS and MERS has been invaluable. This extensive research history has not only provided crucial information about other epidemics but also enabled the development of data-driven techniques. These techniques are pivotal in estimating, modelling, forecasting, and making informed decisions during an epidemic outbreak. This foundation is further strengthened by comprehensive mathematical models for epidemiology and insights into previous coronavirus epidemics.

The solid theoretical foundations of data-driven methods play a critical role. These foundations allow decision-makers to manage pandemics using a variety of tools, each adapted to different contexts and data types. This versatility ensures varying degrees of certainty and efficacy in pandemic management.

Efficient optimization algorithms and solvers also form the strength. Many data-driven approaches in pandemic response are formulated as optimization problems. Addressing these problems efficiently and significantly enhances the speed and effectiveness of responses.

Managing a pandemic requires the gathering and processing of vast amounts of data. The advancements in big data analytics, especially in the 21st century, have provided numerous innovative information systems. These include using mobile phones for contact tracing and employing social networks for real-time epidemic mapping.

The high computation capacity available today significantly aids in combating the pandemic. Continuous advancements in hardware technologies, such as supercomputers, clusters, and cloud computing, equipped with numerous processors and GPUs, have been crucial. This high computational power is essential for developing and running complex data-driven methods effectively.

Together, these strengths underscore the pivotal role of data-driven methods in understanding, managing, and mitigating the impacts of the COVID-19 pandemic, demonstrating the power of technology and data in modern healthcare challenges.

3.1.1.2 Weaknesses

A major weakness identified in the response to COVID-19 was the lack of an interdisciplinary approach among many research groups. The studies and results often did not combine efforts from diverse but crucial fields like epidemiology, data science, system engineering, and economics. This shortfall likely led to biased results and recommendations, as important aspects of public health, advanced statistical tools, dynamic effects, and economic impact were potentially overlooked. The absence of a multi-disciplinary perspective in the research community hindered the ability to formulate a well-rounded and effective response to the pandemic.

Another significant weakness was the slow pace of the academic response. Academic outcomes, primarily in the form of publications and predictive tools, require substantial time for data collection, processing, development, validation, and peer review. In the context of a fast-moving pandemic like COVID-19, this inherent delay in the academic process could impede the timely application and effectiveness of research findings. The need for rapid yet reliable academic responses became critically apparent, highlighting a gap between research timelines and urgent public health requirements.

The initial poor characterization of COVID-19 posed a substantial challenge. Despite extensive research efforts, there remained significant gaps in understanding the virus's characteristics, such as its seasonal behavior and the duration of post-recovery immunity. These uncertainties impaired the capacity of data-driven methods to provide accurate and reliable predictions and strategies for managing the pandemic. The lack of consolidated and comprehensive knowledge about the virus underscored the difficulty in applying data-driven approaches effectively in real-time during the early stages of the pandemic.

These weaknesses underscore the need for a more agile, comprehensive, and interdisciplinary approach to utilizing data-driven methods for pandemic management and response.

3.1.1.3 Opportunities

The opportunities presented by data-driven methodologies in managing the COVID-19 pandemic are vast and critical. These methodologies are applicable across a range of epidemiological sub problems, from estimating epidemic characteristics to

forecasting and assessing the impact of government measures. For example, timely decisions such as partial or total lockdowns, informed by data-driven insights, have been essential in saving thousands of lives while mitigating socio-economic damage.

Furthermore, the coordinated efforts of institutions like the World Health Organization and various Centers for Disease Prevention and Control, both at national and continental levels, have been pivotal. Many governments and research bodies have formed interdisciplinary task forces to develop and implement data-driven strategies against COVID-19. This coordination enhances the effectiveness of the response.

The availability of open data sources has been another significant opportunity. The growing number of institutions and research teams working against COVID-19 has led to an abundance of meaningful, open-data resources. These resources are crucial for addressing the pandemic from a data science perspective, providing a strong foundation for research and analysis.

The substantial social and economic impact of the pandemic has led to considerable public and private funding for related research. This financial support is fostering advancements in data-driven methodologies, enhancing the global capacity to combat the pandemic effectively. This confluence of applicability, coordination, data availability, and funding underscores the potential of data-driven methods in navigating and overcoming the challenges posed by COVID-19.

3.1.1.4 Threats

Statistics play a crucial role in guiding policy decisions, resource allocation, and public health interventions. However, their accuracy and reliability are contingent

on several factors, making it essential to thoroughly assess and address potential threats. One of the primary challenges is reduced government transparency. The reluctance of some governments to fully disclose information pertaining to the pandemic severely restricts access to crucial data. This secrecy leads to a lack of understanding and collaboration, which are essential in managing a global health crisis.

Another major obstacle is the inconsistency of data sets. The data quality is compromised by issues like changing data collection criteria, diverse sources and formats, and non-comparable metrics between countries. This inconsistency hampers the ability to perform accurate analyses and draw reliable conclusions. Additionally, the data is often aggregated without clear timestamps, leading to potential misinterpretations of trends and patterns in the virus's spread.

The complexity of the pandemic also presents difficulties in effectively conveying the results of data-driven analyses to society. The dynamic nature of the pandemic, exacerbated by the interconnectedness of today's world, makes it challenging to communicate findings in a way that is clear and understandable to the general public. This gap in communication leads to misinformation and a lack of proper understanding among the populace.

Furthermore, the urgency to respond to the pandemic has resulted in many published results lacking sufficient validation. In the rush to provide timely recommendations, some findings are released without adequate verification, leading to potentially contradictory or ineffective guidelines. This was evident in the varying advice on mask usage by different health authorities. The absence of rigorous testing and validation of these findings before public dissemination undermines the response

to the pandemic, potentially leading to the adoption of ineffective or even detrimental practices.

Overall, these threats underscore the complexities and challenges in leveraging data-driven methods during the COVID-19 pandemic. Addressing these issues requires collaborative efforts from governments, health organizations, and the scientific community to ensure data transparency, consistency, clarity, and validation.

In summary, the consolidation is as follows:

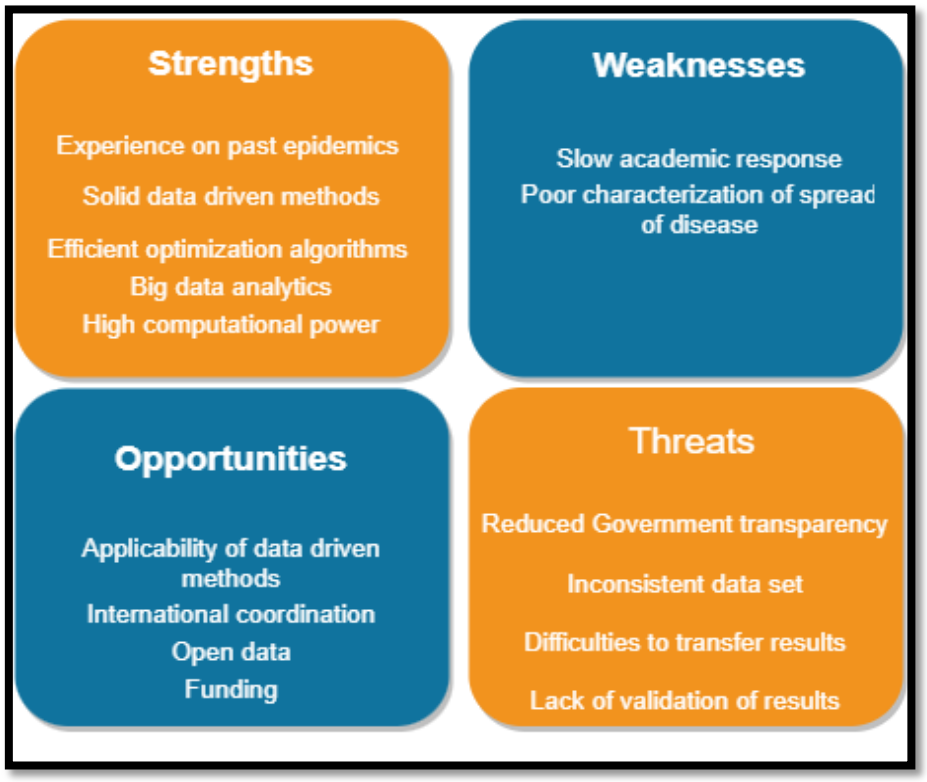


Figure 4: SWOT Analysis

1.1.2 Consolidated Time Series

3.1.2.1 Clustering methods

Table 1: Clusters of States Based on Confirmed, Death, and Recovered Cases Before Lockdown (March 5, 2020 - March 24, 2020)

Cluster 1	Cluster 2	Cluster 3
Maharashtra	Tamil Nadu	Karnataka
Kerala	Bihar	Rajasthan
	Jammu & Kashmir	Punjab
	West Bengal	Delhi
	Madhya Pradesh	Haryana
	Andhra Pradesh	Telangana
		Gujarat
		Uttar Pradesh

During the initial phase of the COVID-19 pandemic in India, spanning from March 5, 2020, to March 24, 2020, states exhibited discernible clustering based on confirmed, death, and recovered cases before the nationwide lockdown was implemented. Table 1 illustrates the clusters during this period, with Cluster 1 solely comprising Maharashtra, and Cluster 2 encompassing several states, including, Tamil Nadu, Bihar, Jammu & Kashmir, West Bengal, Madhya Pradesh, Andhra Pradesh, and Cluster 3 encompassing several states, including Karnataka, Rajasthan, Punjab, Delhi, Haryana, Telangana, Gujarat, and Uttar Pradesh.

Table 2: Clusters of States Based on Confirmed, Death, and Recovered Cases After Lockdown (March 25, 2020 - April 13, 2020)

Cluster 1	Cluster 2	Cluster 3
Maharashtra	Telangana	Haryana
Tamil Nadu	Uttar Pradesh	Bihar
Delhi	Kerala	West Bengal
	Andhra Pradesh	Punjab
	Rajasthan	Karnataka
	Madhya Pradesh	Gujarat
		Jammu & Kashmir

After the initiation of the lockdown from March 25, 2020, to April 13, 2020, a reconfiguration in state clusters occurred, as depicted in Table 2. A cluster dendrogram is drawn to visually represent the discernible clustering of states before and after the nationwide lockdown (Figure 5, Figure 6).

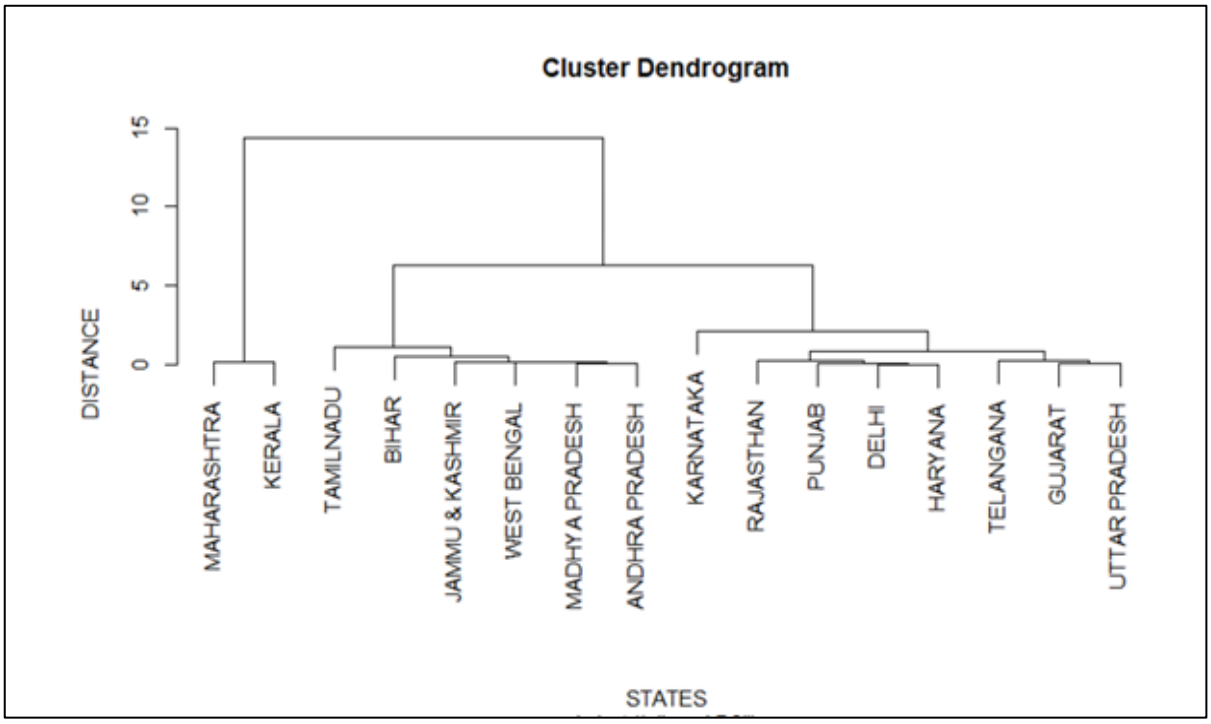


Figure 5: Dendrogram of States Based on Confirmed, Death, and Recovered Cases Before Lockdown (March 5, 2020 - March 24, 2020)

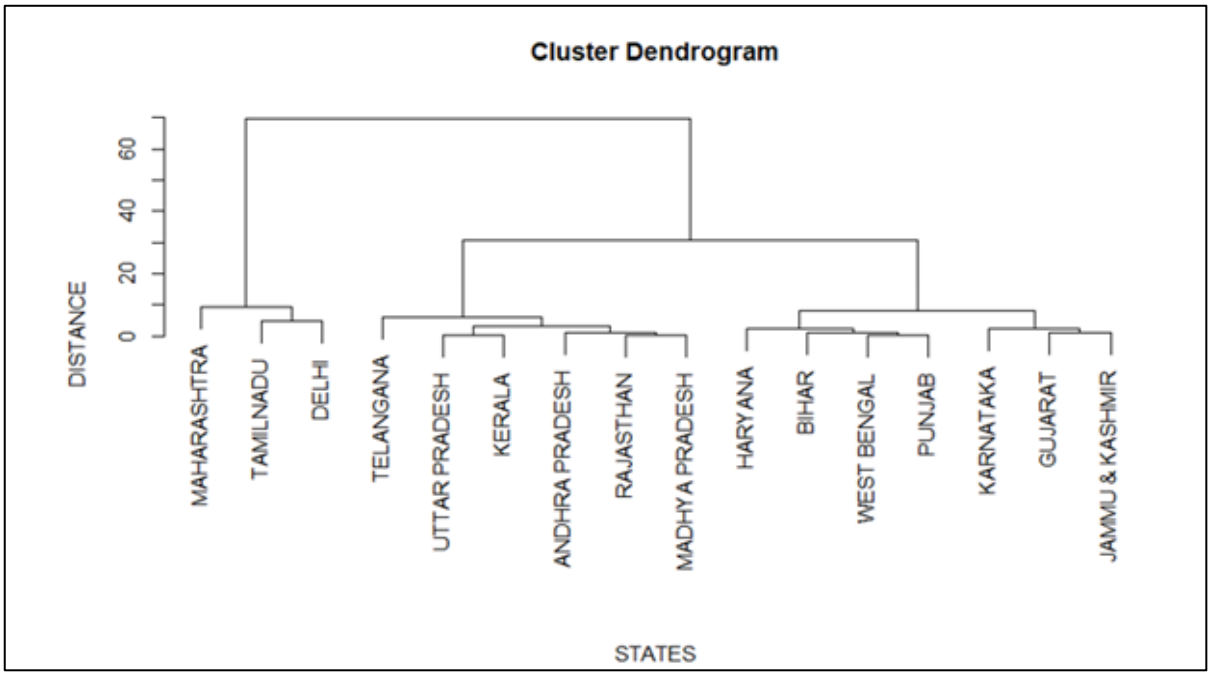


Figure 6: Dendrogram of States Based on Confirmed, Death, and Recovered Cases After Lockdown (March 25, 2020 - April 13, 2020)

Table 3: CFR of selected states and union territories

ID Number	State/Union Territories	CFR as of May 31	CFR as of June 30
1	Andhra Pradesh	1.681	1.295
2	Assam	0.337	0.141
3	Bihar	0.550	0.643
4	Chandigarh	1.384	1.379
5	Chhattisgarh	0.223	0.470
6	Delhi	2.242	3.146
7	Gujarat	6.161	5.720
8	Haryana	1.040	1.632
9	Himachal Pradesh	1.597	0.955
10	Jammu and Kashmir	1.196	1.312
11	Jharkhand	0.888	0.618
12	Karnataka	1.642	1.580
13	Kerala	0.745	0.525
14	Madhya Pradesh	4.346	4.218
15	Maharashtra	3.371	4.479
16	Odisha	0.384	0.335
17	Punjab	1.970	2.547
18	Rajasthan	2.239	2.293
19	Tamil Nadu	0.755	1.323
20	Telangana	3.081	1.643
21	Tripura	0	0.072
22	Uttarakhand	0.667	1.377
23	Uttar Pradesh	2.699	2.943
24	West Bengal	6.023	3.646

Table 3 presents the Case Fatality Rates (CFR) of selected states and union territories in India as of May 31 and June 30. The CFR reflects the percentage of reported deaths in relation to confirmed COVID-19 cases. States like Gujarat, Madhya Pradesh, and

West Bengal exhibit higher CFR values, suggesting a relatively higher fatality rate. In contrast, states such as Assam, Chhattisgarh, and Kerala display lower CFR values, indicating a comparatively lower fatality rate among reported cases.

The table 3 continues to track the CFR for selected states and union territories in India for June 30 also. A comparison with the May 31 data reveals dynamic changes in the CFR values, reflecting the evolving nature of the COVID-19 situation. Notable shifts can be observed in states like Delhi, where the CFR has increased from 2.242 to 3.146, indicating a potential escalation in the severity of the pandemic.

Table 4: Quartiles based on CFR as on 31 May 2020

GROUP 1 (0 to <0.725)	**GROUP 2 (0.725 to <1.490)**	**GROUP 3 (1.490 to <2.356)**	**GROUP 4 (2.356 to 6.161)**
Assam	Chandigarh	Andhra Pradesh	Gujarat
Bihar	Haryana	Delhi	Madhya Pradesh
Chhattisgarh	Jammu and Kashmir	Himachal Pradesh	Maharashtra
Odisha	Jharkhand	Karnataka	Telangana
Tripura	Kerala	Punjab	Uttar Pradesh
Uttarakhand	Tamil Nadu	Rajasthan	West Bengal

Note: Minimum: 0, Quartile 1(25%): 0.725, Median (50%): 1.490, Quartile 3 (75%): 2.356 Maximum: 6.161

Table 5: Quartiles based on CFR as of 30 June 2020

GROUP 1 (0.072 to <0.636)	**GROUP 2 (0.636 to <1.378)**	**GROUP 3 (1.378 to <2.646)**	**GROUP 4 (2.646 to 5.720)**
Assam	Andhra Pradesh	Chandigarh	Delhi
Chhattisgarh	Bihar	Haryana	Gujarat
Jharkhand	Himachal Pradesh	Karnataka	Madhya Pradesh
Kerala	Jammu and Kashmir	Punjab	Maharashtra
Odisha	Tamil Nadu	Rajasthan	Uttar Pradesh
Tripura	Uttarakhand	Telangana	West Bengal

Note: Minimum: 0.072, Quartile 1(25%): 0.636, Median (50%): 1.378, Quartile 3 (75%): 2.646, Maximum: 5.720

Table 4 and Table 5 represent the quartile based splits of CRF of selected states and union territories in India for May 31 and June 30. Severity of the disease increases from Group 1 to Group 4. States under Group 4 are high risk regions with higher severity of disease. These regions should have strict regulations and policies. Higher values of CFR in these states results a limited capacity of the surveillance system to trigger a timely response.

Comparing the two tables, it was seen that some states show a shift from low risk group to high risk group (Bihar, Uttarakhand, Chandigarh, Haryana, and Delhi) and a few shows a shift from high risk group to low risk group (Jharkhand, Kerala, Andhra Pradesh, Himachal Pradesh, and Telangana).

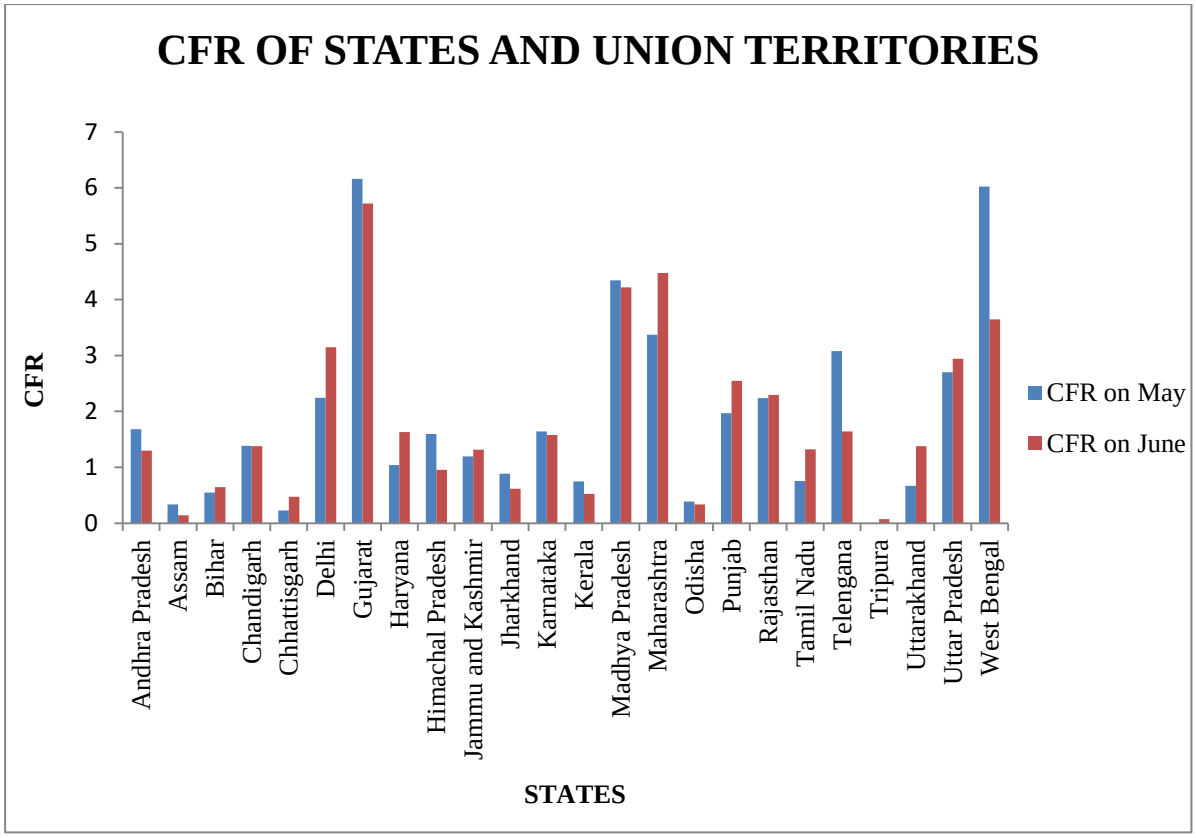


Figure 7: Multiple bar chart showing CFR among different states and union territories for two different time points.

From the graph it is clear that CFR in Bihar, Chhattisgarh, Delhi, Haryana, Jammu and Kashmir, Maharashtra, Rajasthan, Tamil Nadu, Tripura, Uttarakhand, and Utter Pradesh showed an increased severity within a month (Figure 7).

Table 6: RR of selected states and union territories.

ID Number	State/Union Territories	RR as on May 31	RR as on June 30
1	Andhra Pradesh	0.641	0.448
2	Assam	0.137	0.687
3	Bihar	0.444	0.766

4	Chandigarh	0.653	0.802
5	Chhattisgarh	0.228	0.787
6	Delhi	0.435	0.660
7	Gujarat	0.564	0.727
8	Haryana	0.504	0.668
9	Himachal Pradesh	0.354	0.590
10	Jammu and Kashmir	0.387	0.633
11	Jharkhand	0.454	0.760
12	Karnataka	0.341	0.537
13	Kerala	0.475	0.513
14	Madhya Pradesh	0.563	0.762
15	Maharashtra	0.430	0.523
16	Odisha	0.577	0.721
17	Punjab	0.880	0.694
18	Rajasthan	0.666	0.771
19	Tamil Nadu	0.566	0.553
20	Telangana	0.565	0.362
21	Tripura	0.641	0.766
22	Uttarakhand	0.136	0.745
23	Uttar Pradesh	0.592	0.679
24	West Bengal	0.384	0.654

Table 7: Quartiles based on RR as on 31 May 2020

GROUP 1 (0.136 to <0.386)	GROUP 2 (0.386 to <0.490)	GROUP 3 (0.490 to <0.581)	GROUP 4 (0.581 to 0.880)
Assam	Bihar	Gujarat	Andhra Pradesh
Chhattisgarh	Delhi	Haryana	Chandigarh
Himachal Pradesh	Jammu and Kashmir	Madhya Pradesh	Punjab
Karnataka	Jharkhand	Odisha	Rajasthan
Uttarakhand	Kerala	Tamil Nadu	Tripura
West Bengal	Maharashtra	Telangana	Uttar Pradesh

Note: Minimum: 0.136, Quartile 1 (25%): 0.386, Median (50%): 0.490, Quartile 3 (75%): 0.581, Maximum: 0.880

Table 8: Quartiles based on RR as on 30 June 2020

GROUP 1 (0.362 to <0.581)	GROUP 2 (0.581 to <0.683)	GROUP 3 (0.683 to <0.761)	GROUP 4 (0.761 to 0.802)
Andhra Pradesh	Delhi	Assam	Bihar
Karnataka	Haryana	Gujarat	Chandigarh
Kerala	Himachal Pradesh	Jharkhand	Chhattisgarh
Maharashtra	Jammu and Kashmir	Odisha	Madhya Pradesh
Tamil Nadu	Uttar Pradesh	Punjab	Rajasthan
Telangana	West Bengal	Uttarakhand	Tripura

Note: Minimum: 0.362, Quartile 1 (25%): 0.581, Median (50%): 0.683, Quartile 3 (75%): 0.761, Maximum: 0.802

Table 6 presents recovery rate (RR) of selected states and union territories in India as of May 31 and June 30. Table 7 and Table 8 represent the quartile-based splits of RR of selected states and union territories in India for May 31 and June 30. Recovery rate among states increases from Group 1 to Group 4. Punjab has the highest recovery rate per death of 44 as on May 30, 2020 while Uttarakhand, with 5 deaths has the lowest recovery rate. By June 30, 2020 Chandigarh shows a high recovery rate while Telangana has the lowest recovery rate (Figure 8).

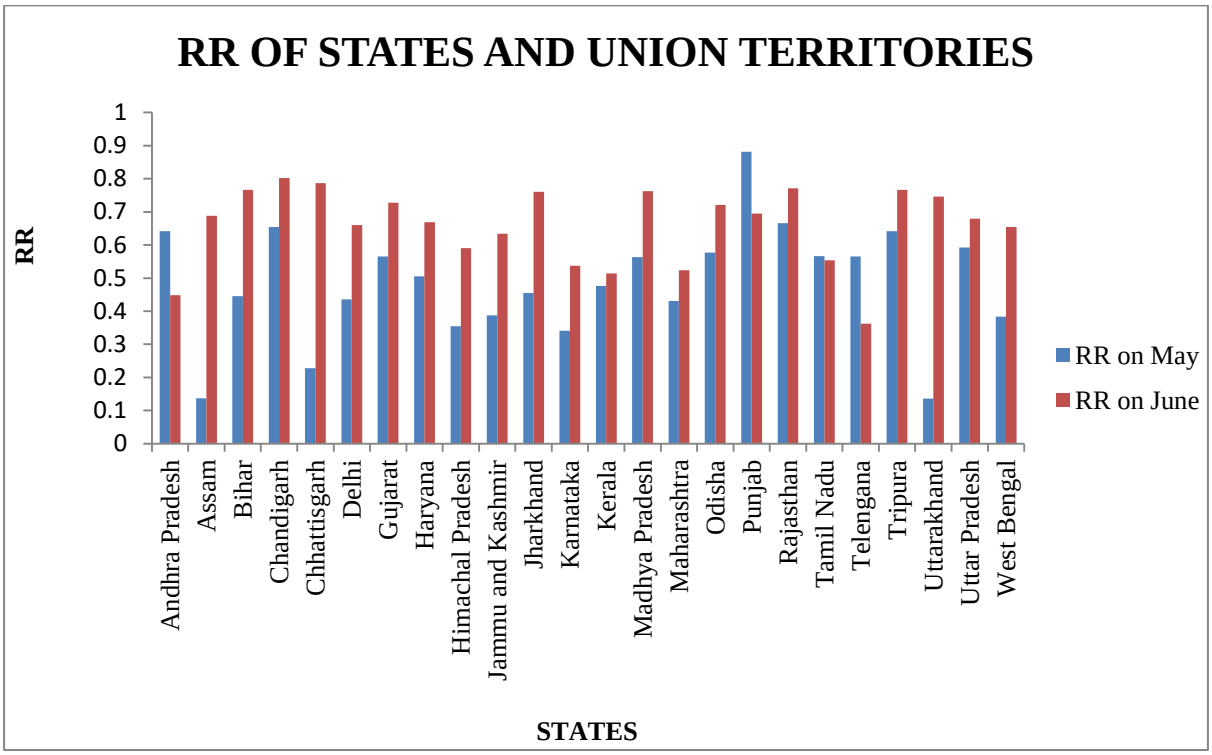


Figure 8: Multiple bar chart showing RR among different states and union territories for two different time points.

The figure indicates that there is a wide variance between recovery rates of states within a month. Recovery rate for most of the states have increased expect Andhra Pradesh, Punjab, and Telangana (Figure 8).

Table 9: MR of selected states and union territories.

ID Number	State/Union Territories	MR as on May 31	MR as on June 30
1	Andhra Pradesh	0.121	0.363
2	Assam	0.012	0.035
3	Bihar	0.019	0.060
4	Chandigarh	0.378	0.568
5	Chhattisgarh	0.003	0.051
6	Delhi	2.477	15.964
7	Gujarat	1.666	3.023
8	Haryana	0.078	0.915
9	Himachal Pradesh	0.072	0.131
10	Jammu and Kashmir	0.228	0.774
11	Jharkhand	0.015	0.045
12	Karnataka	0.078	0.370
13	Kerala	0.026	0.066
14	Madhya Pradesh	0.472	0.777
15	Maharashtra	1.955	6.772
16	Odisha	0.016	0.055
17	Punjab	0.158	0.497
18	Rajasthan	0.281	0.591
19	Tamil Nadu	0.221	1.581
20	Telangana	0.219	0.723
21	Tripura	0	0.027
22	Uttarakhand	0.049	0.387
23	Uttar Pradesh	0.101	0.336
24	West Bengal	0.338	0.715

Table 10: Quartiles based on MR as on 31 May 2020

GROUP 1 (0 to <0.025)	GROUP 2 (0.025 to <0.110)	GROUP 3 (0.110 to <0.295)	GROUP 4 (0.295 to 2.477)
Assam	Haryana	Andhra Pradesh	Chandigarh
Bihar	Himachal Pradesh	Jammu and Kashmir	Delhi
Chhattisgarh	Karnataka	Punjab	Gujarat
Jharkhand	Kerala	Rajasthan	Madhya Pradesh
Odisha	Uttarakhand	Tamil Nadu	Maharashtra
Tripura	Uttar Pradesh	Telangana	West Bengal

Note: Minimum: 0 Quartile 1 (25%): 0.025 Median (50%): 0.110 Quartile 3 (75%): 0.295 Maximum: 2.477

Table 11: Quartiles based on MR as on 30 June 2020

GROUP 1 (0.027 to <0.064)	GROUP 2 (0.064 to <0.442)	GROUP 3 (0.442 to <0.775)	GROUP 4 (0.774 to 15.963)
Assam	Andhra Pradesh	Chandigarh	Delhi
Bihar	Himachal Pradesh	Jammu and Kashmir	Gujarat
Chhattisgarh	Karnataka	Punjab	Haryana
Jharkhand	Kerala	Rajasthan	Madhya Pradesh
Odisha	Uttarakhand	Telangana	Maharashtra
Tripura	Uttar Pradesh	West Bengal	Tamil Nadu

Note: Minimum: 0.027 Quartile 1 (25%): 0.064 Median (50%): 0.442 Quartile 3 (75%): 0.775 Maximum: 15.963

Table 9 presents mortality rate (MR) of selected states and union territories in India as of May 31 and June 30. Table 10 and Table 11 represent the quartile-based splits of MR of selected states and union territories in India for May 31 and June 30. The mortality rate for states increases from Group 1 to Group 4. High mortality rate indicate increasing number of deaths due to COVID-19 pandemic. Even though the magnitude of mortality rate has changed, most of the states are likely to remain in their respective groups except Haryana, Andhra Pradesh, Madhya Pradesh, Chandigarh, Tamil Nadu and Telangana (Figure 9).

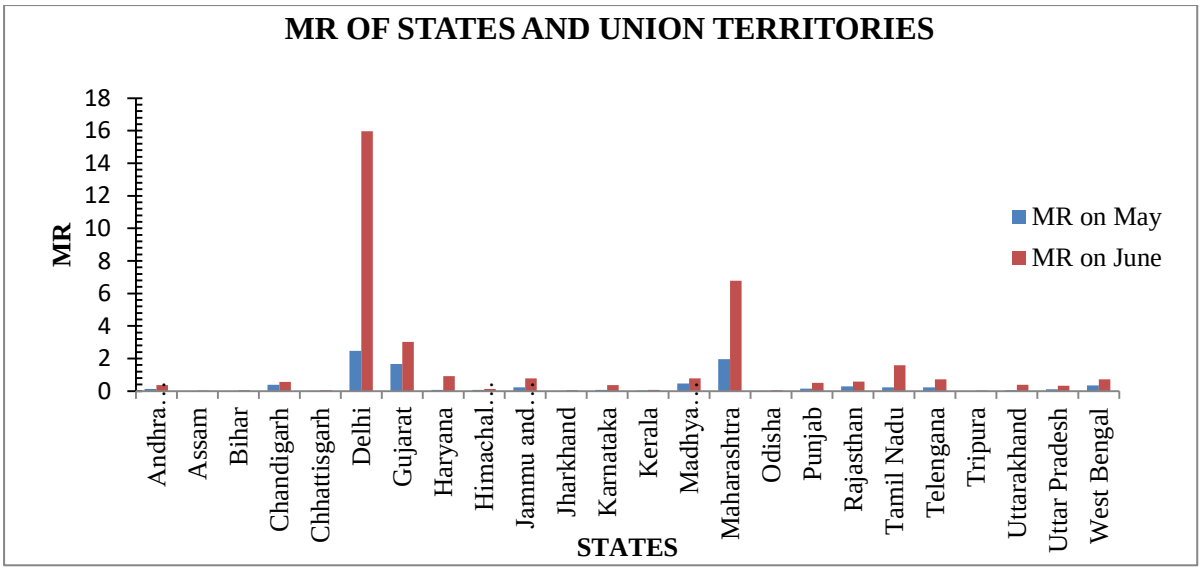


Figure 9: Multiple bar chart showing MR among different states and union territories for two different time points.

Delhi and Maharashtra show a hike in mortality rate within a month. The mortality rate for Delhi increased from 2.477 to 15.964 because the death due to COVID-19 increased from 416 to 2680. In case of Maharashtra the mortality rate increased from 1.955 to 6.772 with number of deaths from 2197 to 7610 (Figure 9).

1.1.2.2 Time-series theory

Table 12: Augmented Dickey-Fuller test of the time series data

Dickey-Fuller	Lag order	Significance
1.30	4	0.99
alternative hypothesis: stationary		

Table 12 displays the Augmented Dickey-Fuller test results for time-series data, indicating a Dickey-Fuller Statistic of 1.30 with a lag order of 4 and a significance level of 0.99. The outcomes support the alternative hypothesis of stationarity, suggesting that the analyzed data may exhibit a stationary behavior.

Table 13: In-sample Forecasting of COVID-19 Confirmed Cases

Days	Forecasted Value of Confirmed Cases	True Value of Confirmed Cases	Error
20-05-2020	5694	5720	26
21-05-2020	5903	6023	120
22-05-2020	6496	6536	40
23-05-2020	6640	6665	25
24-05-2020	6979	7111	132

Table 13 in the thesis analysis displays predicted and actual COVID-19 confirmed cases for specific dates along with the error terms, which quantifies the variance between the forecasted and true values.

Table 14: Ljung Box Test

χ^2	df	Significance
0.27	1	0.60

Table 14 reports the outcomes of the Ljung-Box test, a statistical diagnostic for assessing autocorrelation in time series data. The table enumerates the test statistic (χ^2), the degrees of freedom (df), and the associated significance level. The computed test statistic is 0.27 with 1 degree of freedom, and the p-value, or significance level, is recorded as 0.60. Under the chosen significance threshold of 0.60, there exists insufficient empirical support to reject the null hypothesis. Therefore, there is not statistically significant evidence to assert the presence of autocorrelation within the analyzed time series data.

Table 15: Forecasting of COVID-19 Confirmed Cases, India (May 31, 2020 - June 24, 2020)

Days	**Confirmed Cases**	**Lower Bound 80% CI**	**Upper Bound 80% CI**	**Lower Bound 95% CI**	**Upper Bound 95% CI**
31-05-2020	8893	8242	9544	7898	9888
06-06-2020	10781	9599	11963	8974	12588
13-06-2020	12669	10848	14490	9884	15454
24-06-2020	15097	12324	17870	10855	19338

Table 15 summarizes COVID-19 forecasts for India from May 31, 2020, to June 24, 2020. It includes predicted cases for specific dates, along with 80% and 95% confidence intervals, indicating the range of uncertainty.

Table 16: Interrupted Time Series of COVID-19 Confirmed Cases, India (January 30, 2020 - May 31, 2020)

Coefficients	Estimate	P-value
(Intercept)	0.04	0.700
Lockdown impact	0.70	0.001*

Note: *p < 0.001

Time Series analysis for COVID-19 confirmed cases in India from January 30, 2020, to May 31, 2020 is presented in Table 16. The table displays coefficients, their estimates, and associated p-values. The lockdown impact coefficient, representing the effect of the lockdown on confirmed cases, has an estimate of 0.70 with a highly significant p-value of 0.001. This succinctly captures the key findings of the Interrupted Time Series analysis, emphasizing the significant impact of the lockdown on COVID-19 confirmed cases in India within the specified timeframe.

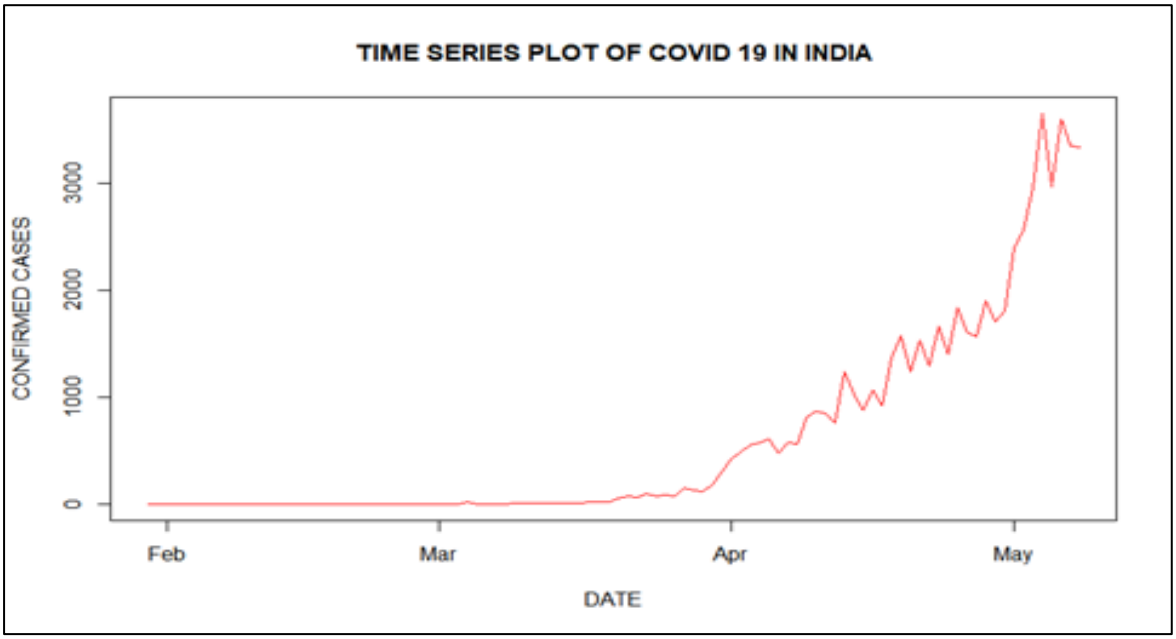


Figure 10: Time Series Plot of COVID-19 Confirmed Cases, India (January 30, 2020 - May 31, 2020)

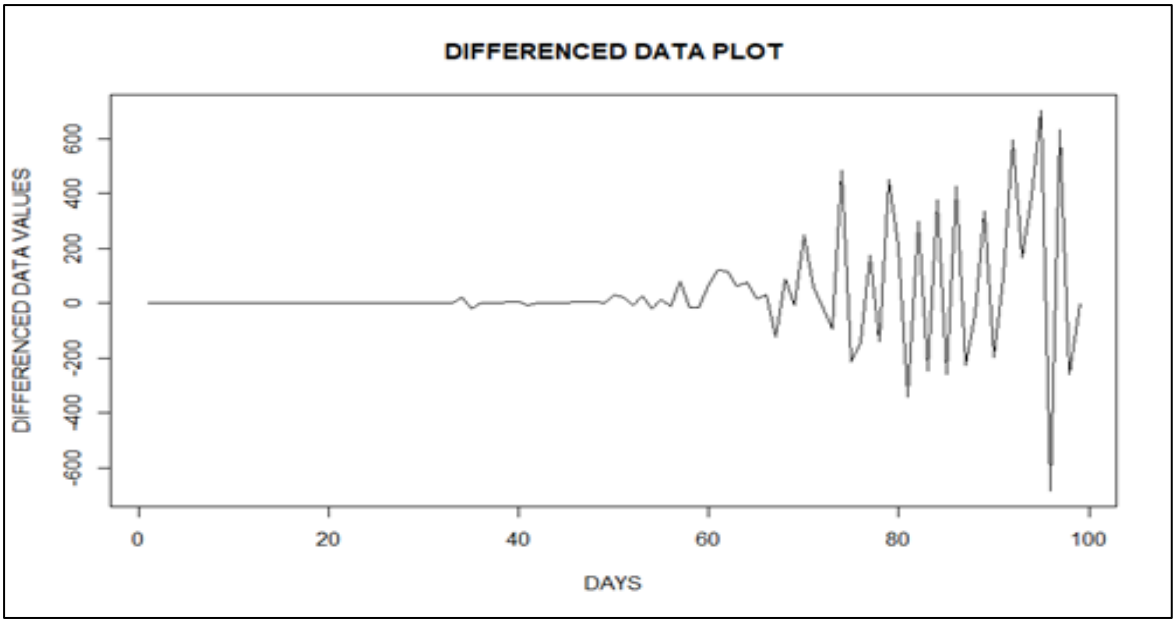


Figure 11: Plot of the Differenced Data of COVID-19 Confirmed Cases, India (January 30, 2020 - May 31, 2020)

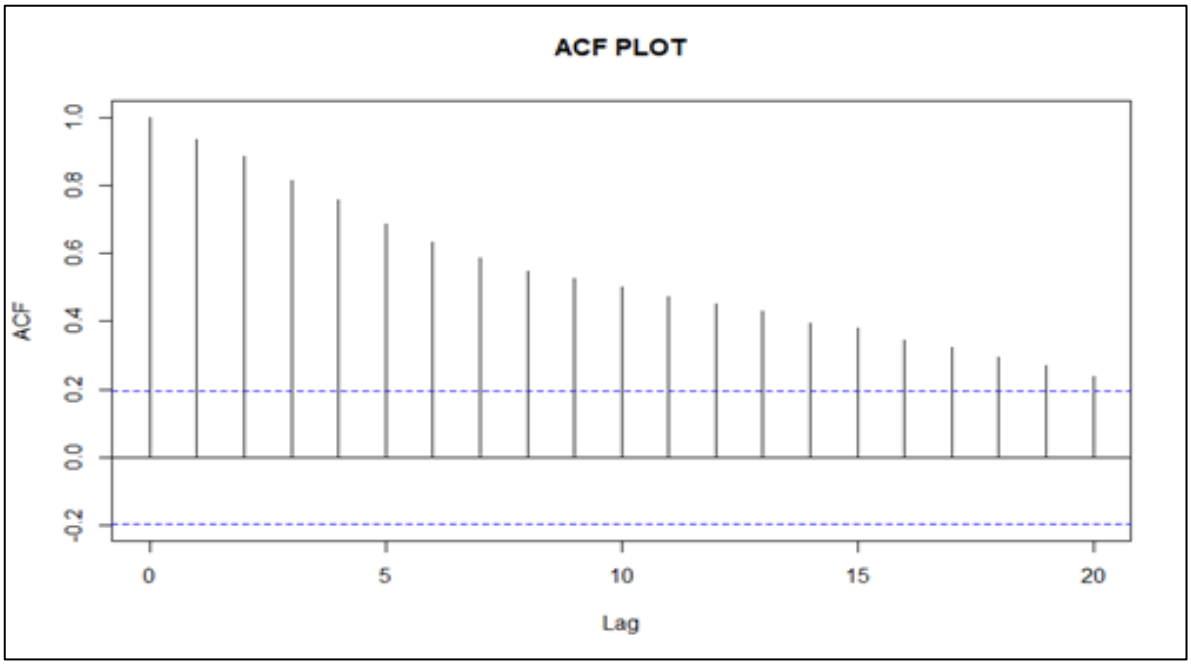


Figure 12: ACF Plot of COVID-19 Confirmed Cases, India (January 30, 2020 - May 31, 2020)

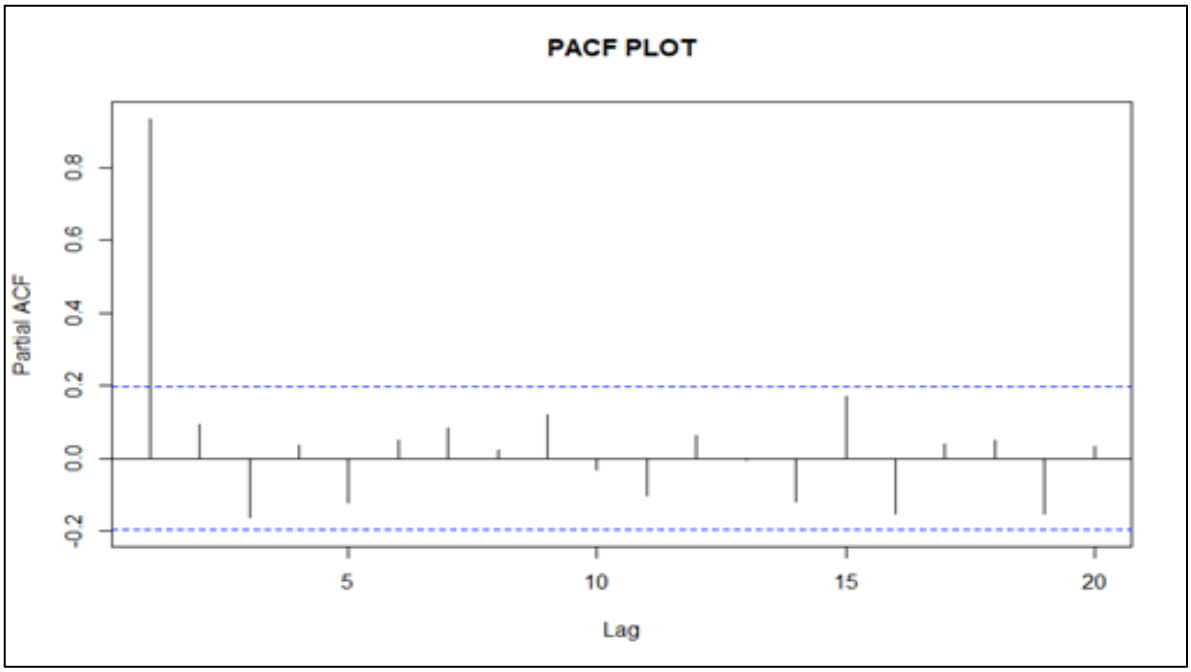


Figure 13: PACF Plot of COVID-19 Confirmed Cases, India (January 30, 2020 - May 31, 2020)

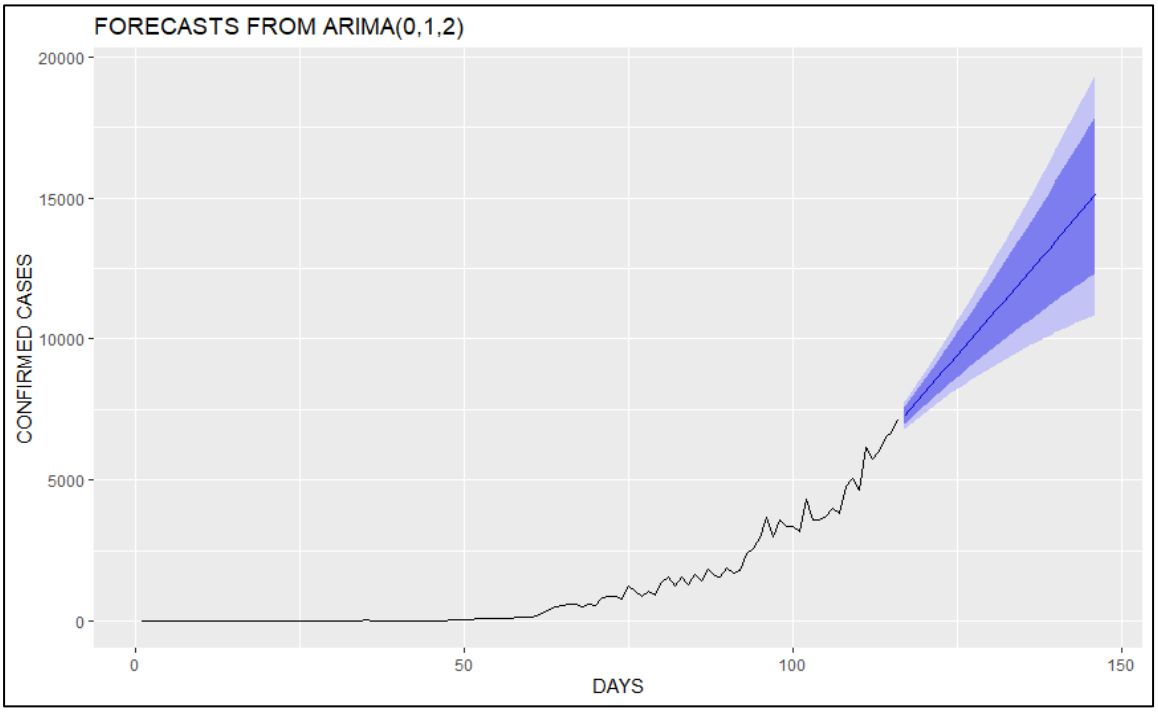


Figure 14: Plot of Forecast of COVID-19 Confirmed Cases, India (May 31, 2020 - June 23, 2020)

Graphical presentations for the time-series data are presented above. Figure 10 presents a time series plot illustrating the progression of COVID-19 confirmed cases in India from January 30, 2020, to May 31, 2020. In Figure 11, the plot displays the differenced data of COVID-19 confirmed cases for the same period. Figure 12 presents an Autocorrelation Function (ACF) plot for COVID-19 confirmed cases in India from January 30, 2020, to May 31, 2020. This plot gives the correlation between each data point and its lagged values, providing insights into the temporal patterns and potential dependencies in the confirmed cases data. Figure 13 displays a Partial Autocorrelation Function (PACF) plot for the same period. The PACF plot aids in identifying the direct relationship between data points while removing the influence of intermediate lagged values. Figure 14 showcases a plot of the forecasted COVID-19 confirmed cases in India from May 31, 2020, to June 23, 2020. This represents a glimpse into the anticipated trajectory of confirmed cases, providing valuable information for understanding the expected trends in the near future.

1.1.3 Estimation of the state of the pandemic

1.1.3.1 Real-Time Epidemiology

The exponential growth fitting results are depicted. It shows that the early outbreak of COVID-19 data in India was largely following an exponential growth at an estimated rate of 1.127 per day. The estimated coefficients and 95% CI for this model. Depicts the coefficient of determination R^2 = 0.963 for this fit. In India exponential model forecasts considerable uncertainties between the actual and predicted number of cases at the later phase. The rapid increase in predicted cases of exponential model due to the inability of this model to account for eventual decay in an epidemic (Figure15).

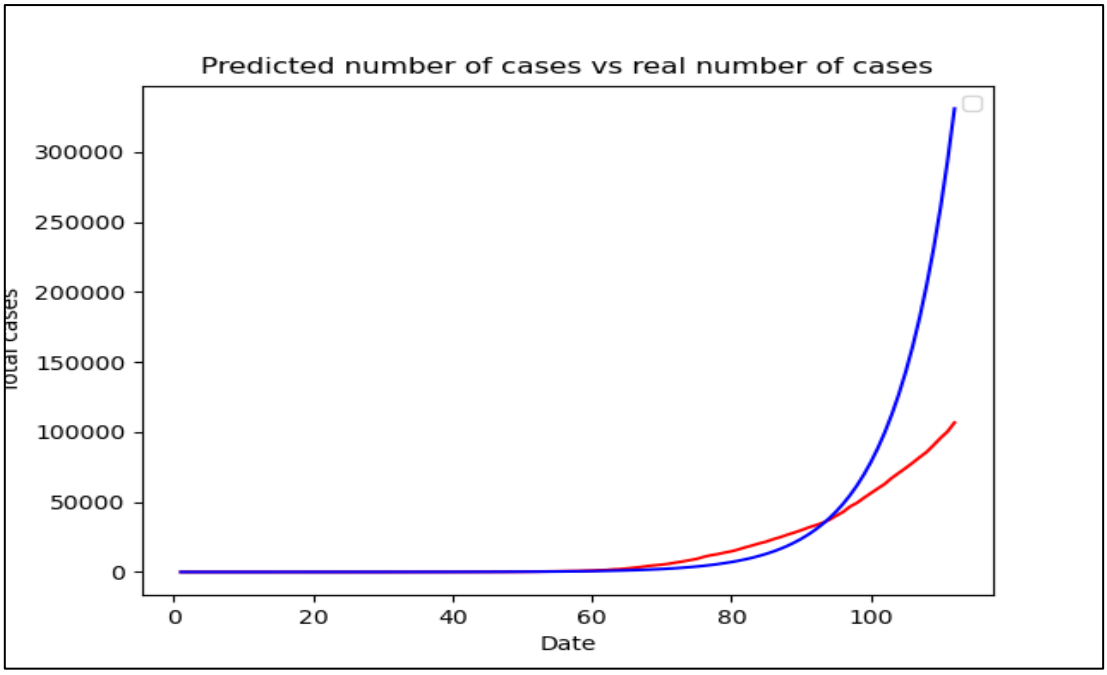


Figure 15: Predictions using the exponential growth model with respect to the number of original cases.

Table 17: Estimated coefficients using the Exponential Growth model

	Coefficient	**Standard error**	**T**	**P>|t|**	**[0.025**	**0.975]**
Constant	-0.71	0.14	-4.91	0.000	-1.09	-0.42
Time	0.11	0.002	53.34	0.000	0.11	0.12

Table 18: Statistics for Exponential Growth Model

R-squared	**Adjusted R-squared**	**F-Statistic**	**AIC**	**BIC**
0.963	0.96	2845	261.0	266.5

The SIR model is developed for predicting the growth of COVID-19 in India using the basic reproduction number. Table 17 presents the estimated coefficients using the Exponential Growth model. The population size N is taken as 1.3×10^9. The model parameter values of the epidemic models are used in this study to estimate R_0 as it determines how the number of infections grow and spread in the population (Table 18).

The study calibrates the parameters of the SIR model to fit the reported data of confirmed cases till May 20. The plot of model fitting using the SIR model is depicted in Figure 16. Basic reproduction number (R_0) is estimated to be 1.012 which means a single individual was infected 1.012 other individuals.

The estimated value of R_0 from the SIR model suggests that disease will spread and eventually invade in India. The decrease in R_o value due to strict lockdown measures implemented by India at the earlier phase of the pandemic suggesting if lockdown is weakened this will increase rapidly. Table 19 shows the actual and predicted number of infections using the SIR model. From this table, the discrepancies between actual and predicted cases using the SIR model by fitting the model to the official reported data was inferred.

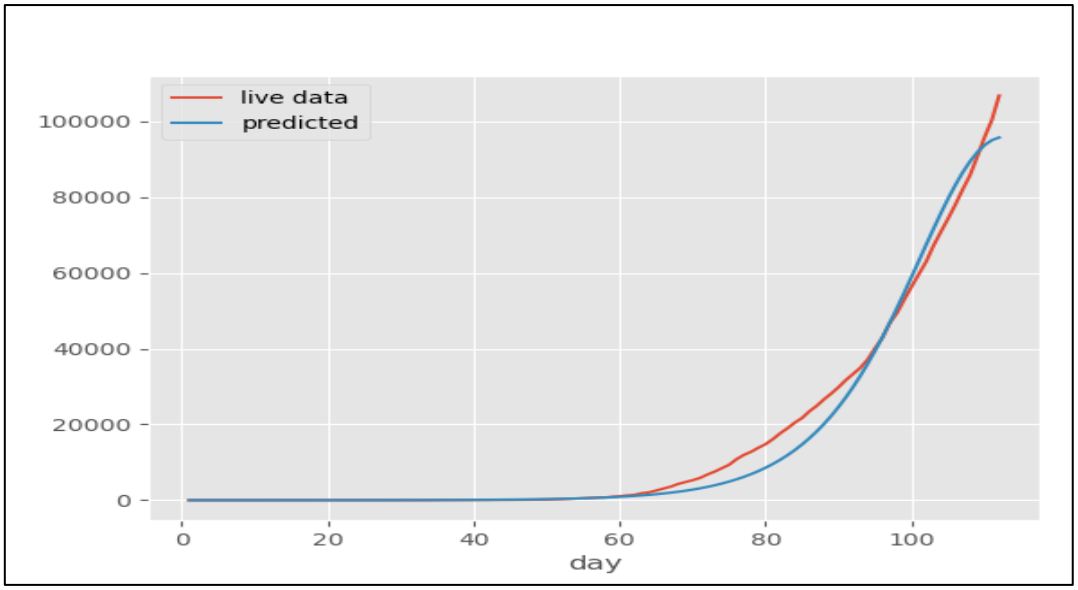


Figure 16: Fit of SIR model to reported data in India.

Table 19: Predicted and actual values of confirmed cases using fitted SIR model

SIR model		
Day	**Live Data**	**Predicted**
1	1.0	1.00
2	1.0	1.12
3	2.0	1.25
4	3.0	1.41
5	3.0	1.58
..	…	…
108	85710.0	89365.39
109	90637.0	91786.86
110	95759.0	93680.07
111	100325.0	94998.06
112	106750.0	95707.45

The study also validates the current spreading rate using the SIRD model. The least-square fitting of the real data using the SIRD model gives us the basic reproduction number as 1.1129. Thus, the number of infections in the population grows and spreads. Figure 17 illustrates the plot of the fitted SIRD model. Predicted and actual values of infection in India using the SIRD model are depicted (Table 19).

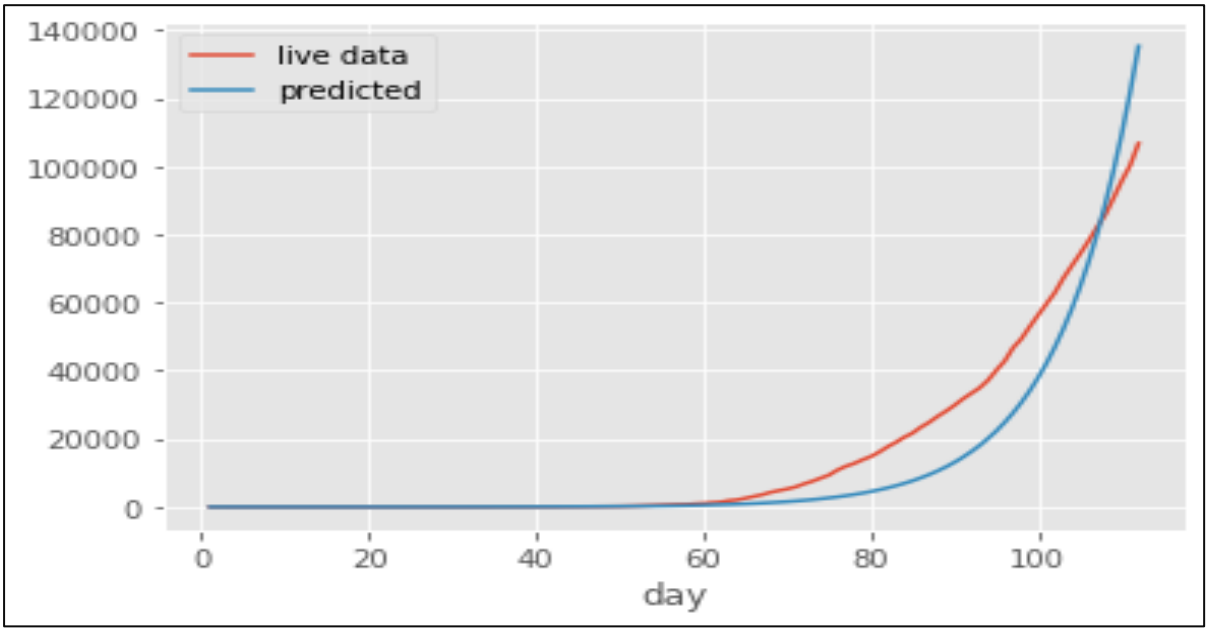


Figure 17: Fit of SIRD model to reported data in India.

Table 20: Predicted and actual values of confirmed cases using fitted SIRD model.

SIRD model		
Day	**Live data**	**Predicted**
1	1.0	1.00
2	1.0	1.11
3	2.0	1.23
4	3.0	1.37
5	3.0	1.53
..	…	…

108	85710.0	88674.57
109	90637.0	98571.50
110	95759.0	109564.12
111	100325.0	121771.64
112	106750.0	135325.74

According to the results of both the models, the basic reproduction number of COVID-19 in India is estimated to be 1.012 and 1.1129 for SIR and SIRD models respectively. The basic reproduction number obtained by two models has only slight variations. However, both models show that the virus spreads through the population increases the number of infected people.

1.1.3.2 Epidemic Wave Insight

The prediction model was performed with respect to each state. Histogram for the standardized residual of the fitted regression model with log of active cases as dependent variable looked plausible where residuals approximated a normal distribution with no weird shapes or low values or extremely high values. Normal P-P plot demonstrated that the distribution of standardized residual deviates only moderately from a distribution that was classically bell-shaped.

The scatter plot indicated that the distribution of residuals across the expected values was equal, suggesting that the model was homoscedastic. Statistical models have been constructed for the duration of patients under COVID-19 across all states in India. The model fitted well for all the regions. Log of New Cases and log of observed cases was plotted against the log of Predicted Active Cases with the corresponding upper and lower confidence limits.

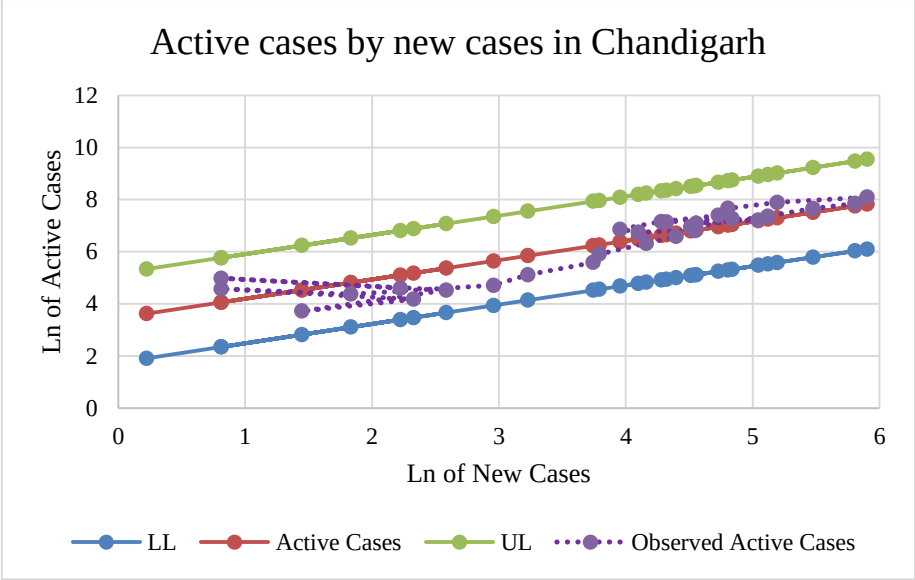


Figure 18: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Chandigarh

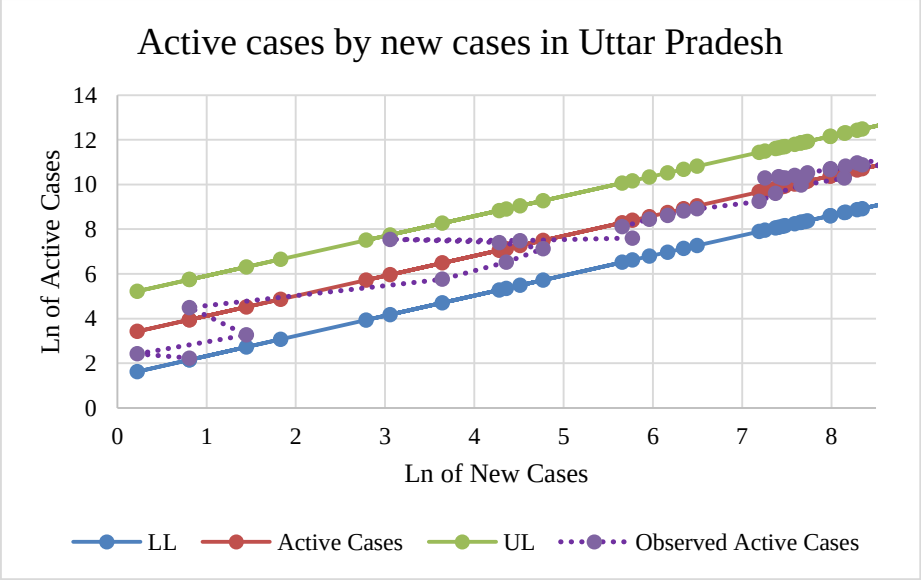


Figure 19: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Uttar Pradesh

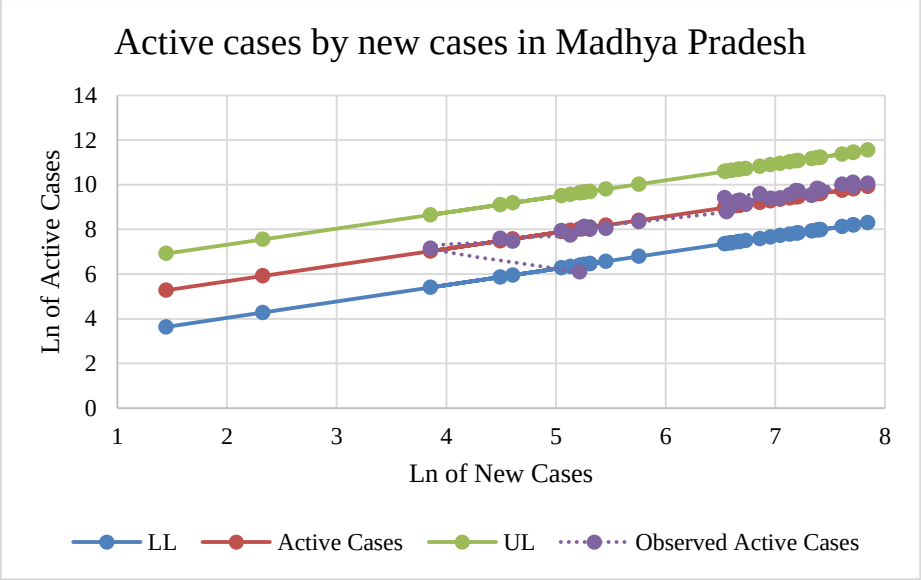


Figure 20: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Madhya Pradesh

Log of New Cases was plotted against log of predicted active cases with the corresponding upper and lower confidence limits and was found linear for central states of India comprising the states Chhattisgarh (Figure 18), Uttar Pradesh (Figure 19), and Madhya Pradesh (Figure 20). Even though the model fit was enough to predict the average duration of COVID-19, there was a significant difference in the average duration between central states.

Madhya Pradesh(R^2=0.76) has an average duration as high as 68 days (CI= 50.348-92.814), whereas Uttar Pradesh(R^2=0.88) with 25 days (CI = 19.353 - 32.484) and Chhattisgarh(R^2=0.94) with 12 days (CI = 10.406 - 14.583).

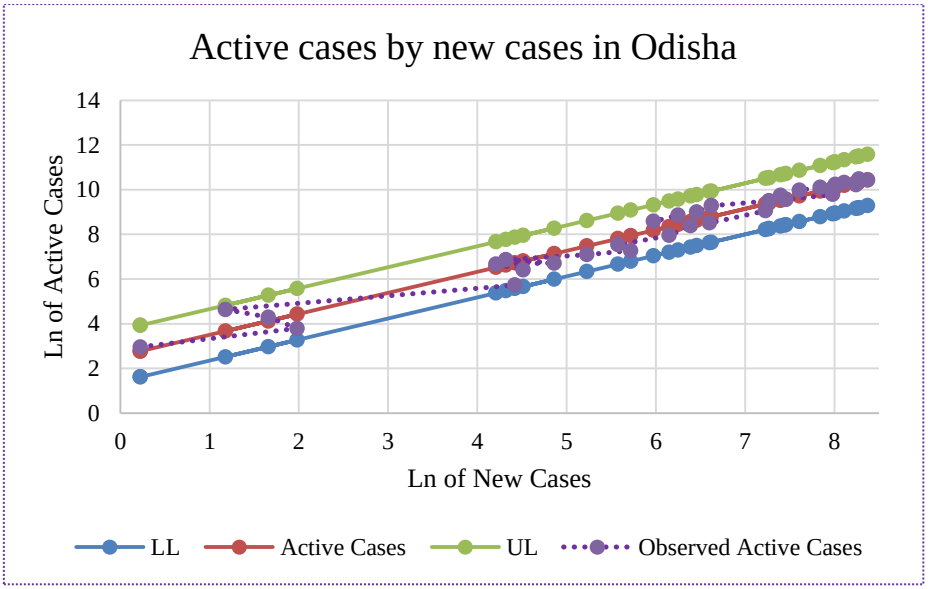


Figure 21: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Odisha

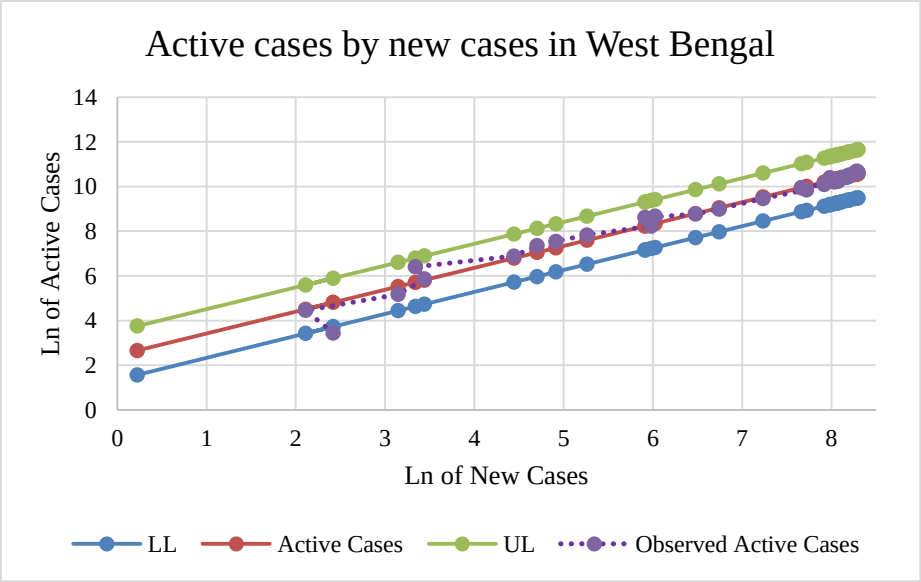


Figure 22: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in West Bengal

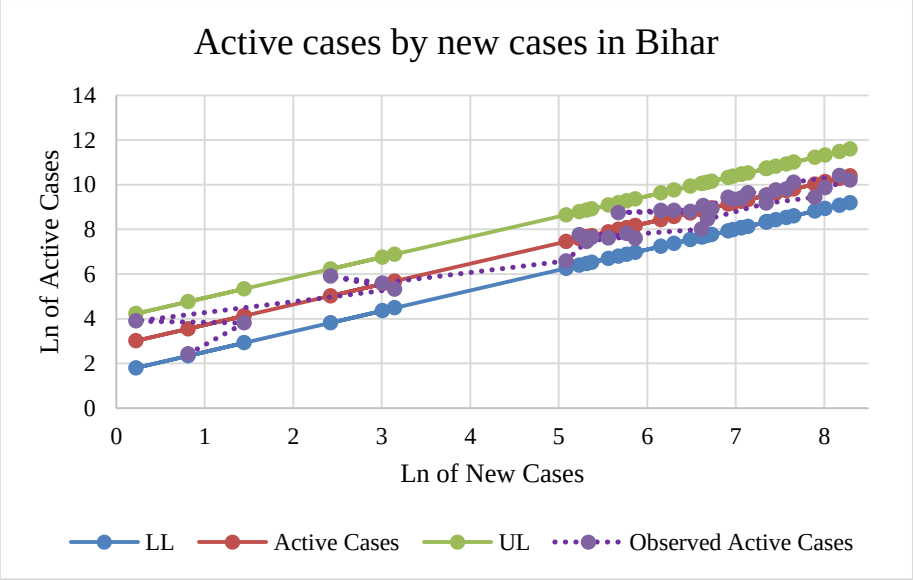


Figure 23: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Bihar

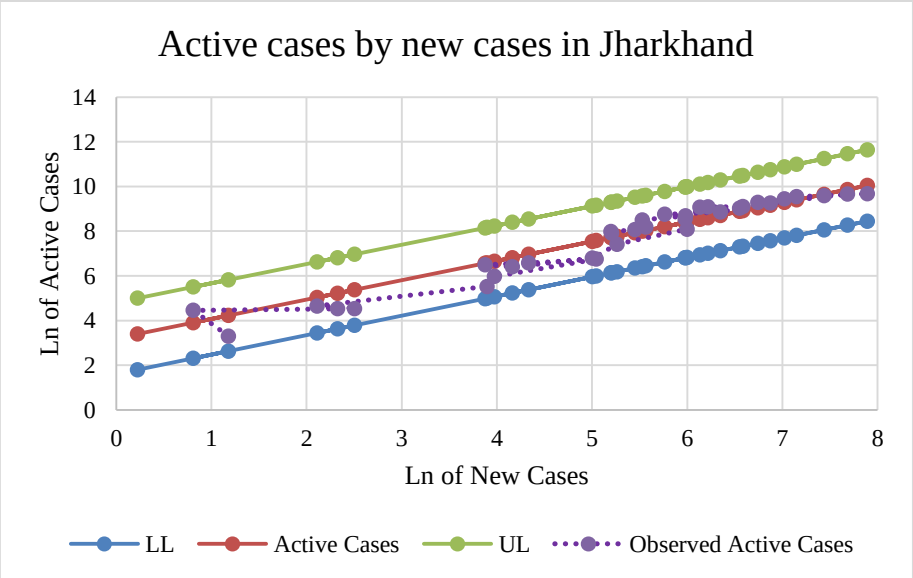


Figure 24: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Jharkhand

In the eastern region, log of new cases against log of predicted active cases with the upper and lower confidence limits was plotted for Odisha (Figure 21), West Bengal (Figure 22), Bihar (Figure 23), and Jharkhand (Figure 24). The average duration in days where a patient remains a patient is 11 (CI = 9.497-13.881) in West Bengal (R^2=0.94) while the average days' rise to 25 days (CI = 19.806-30.632) in Jharkhand (R^2=0.87). Duration days of disease in Odisha(R^2=0.95) and Bihar(R^2=0.92) are 13 days (CI = 11.168-15.035) and 17 days (CI = 13.637-20.132).

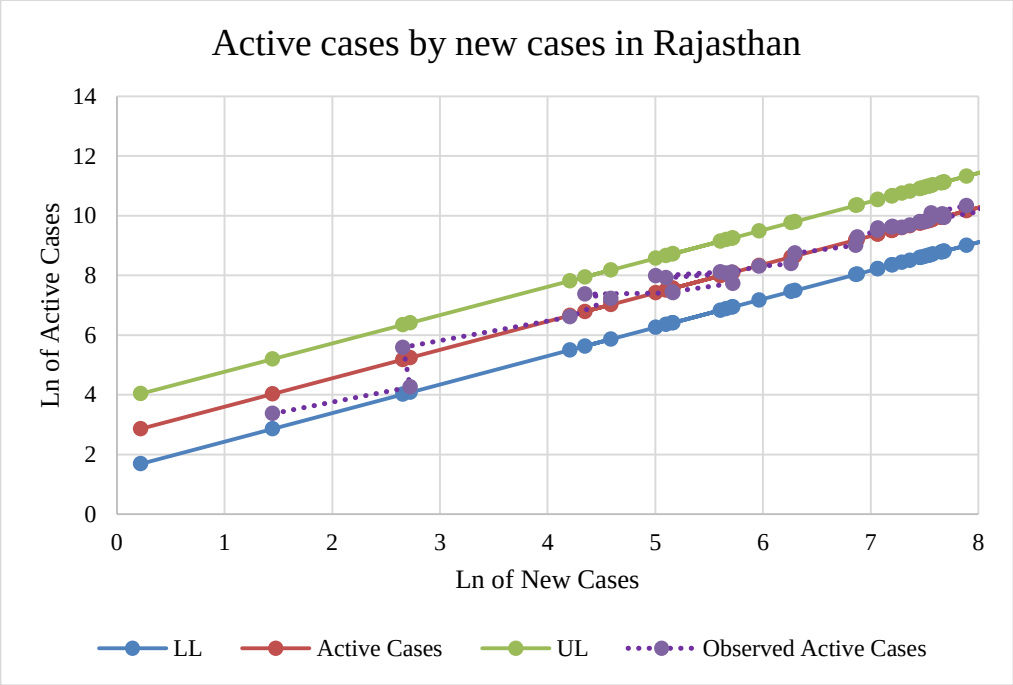


Figure 25: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Rajasthan

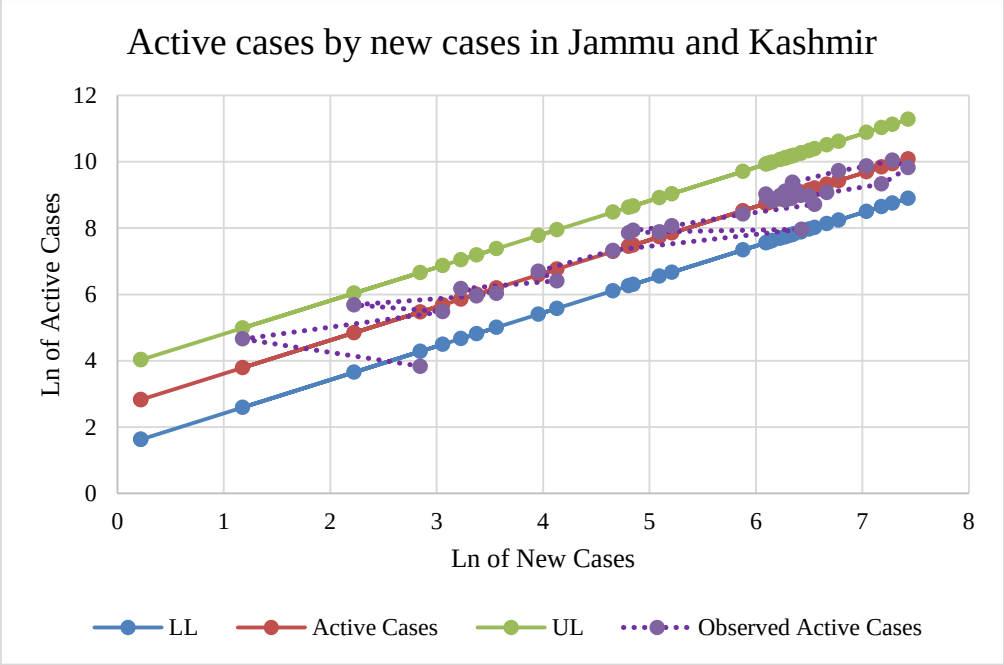


Figure 26: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Jammu Kashmir

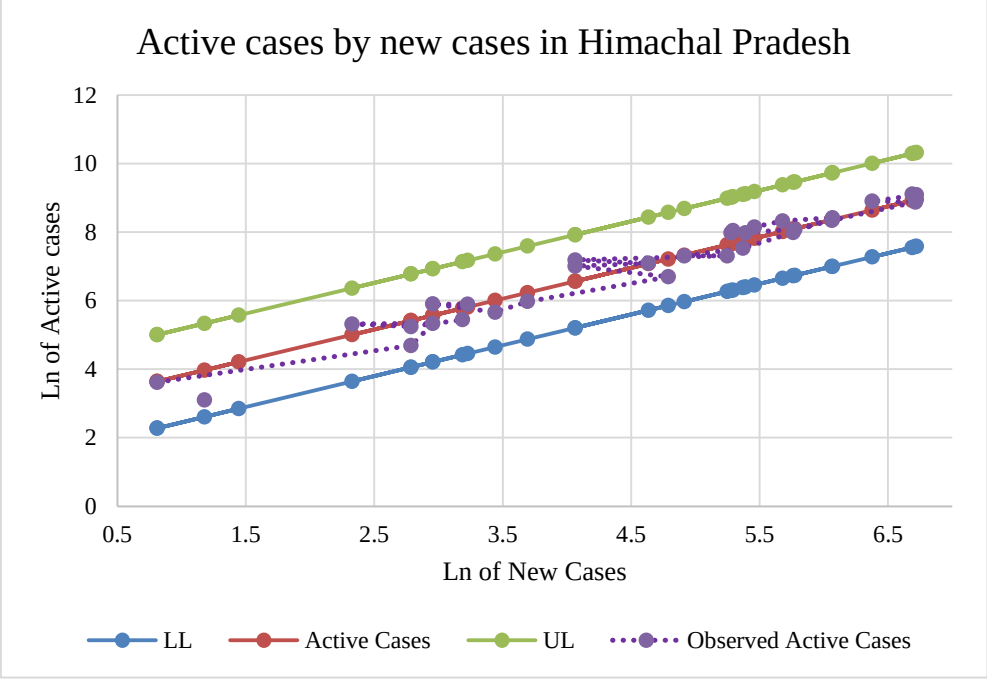


Figure 27: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Himachal Pradesh

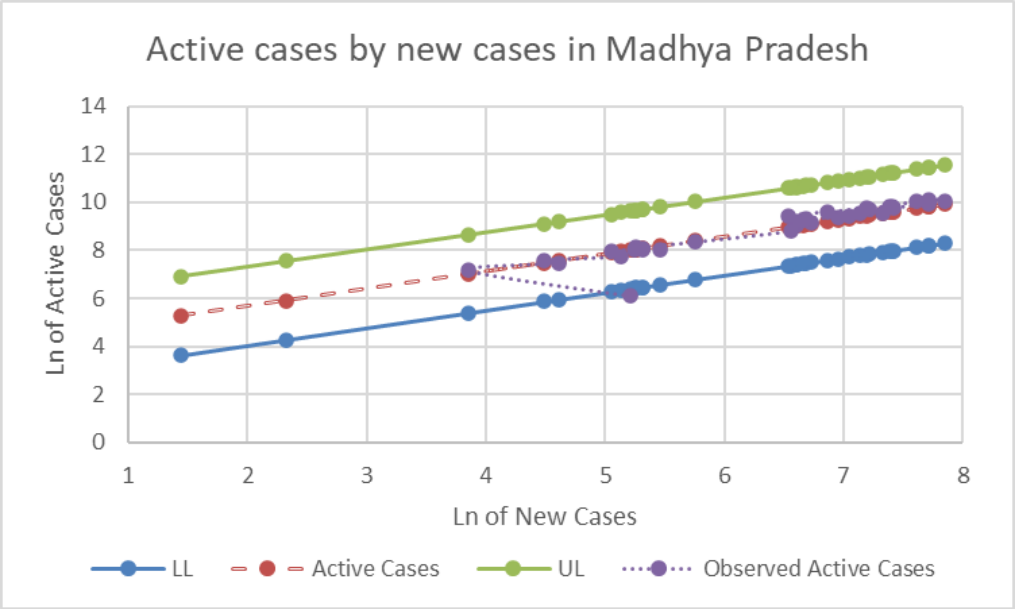


Figure 27: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Madhya Pradesh

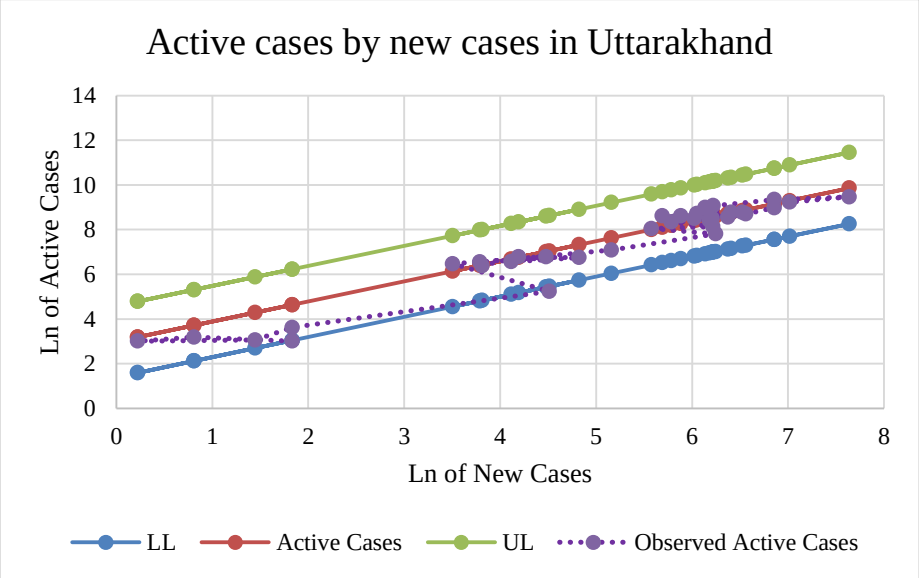


Figure 28: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Uttarakhand

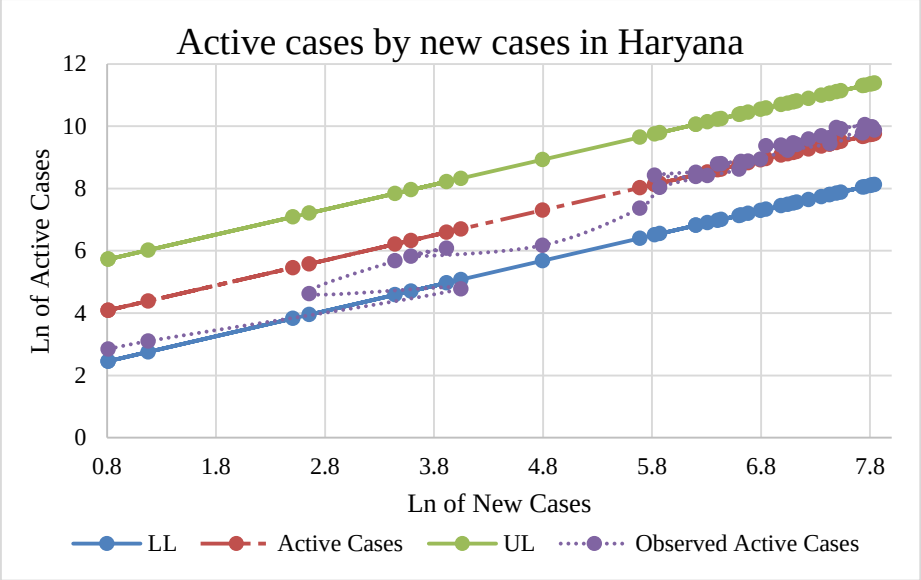


Figure 29: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Haryana

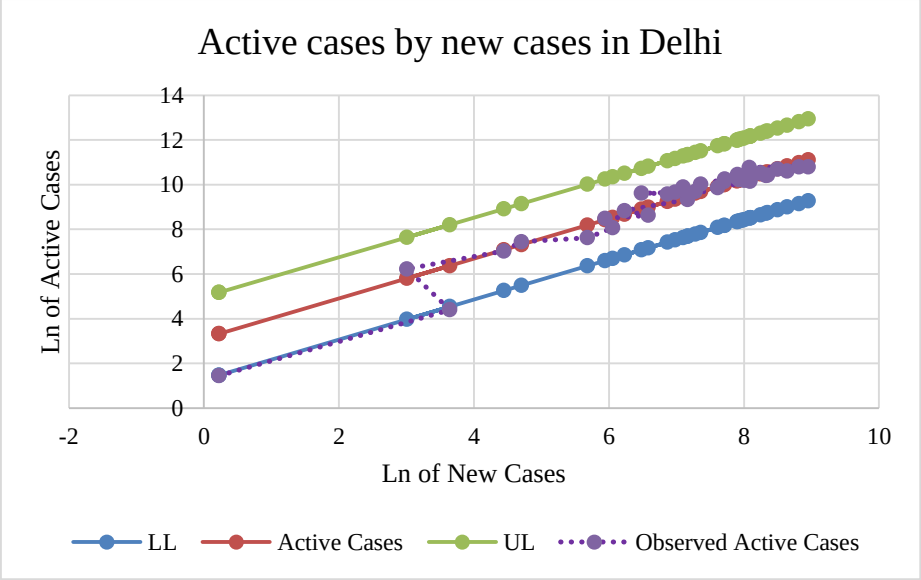


Figure 30: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Delhi

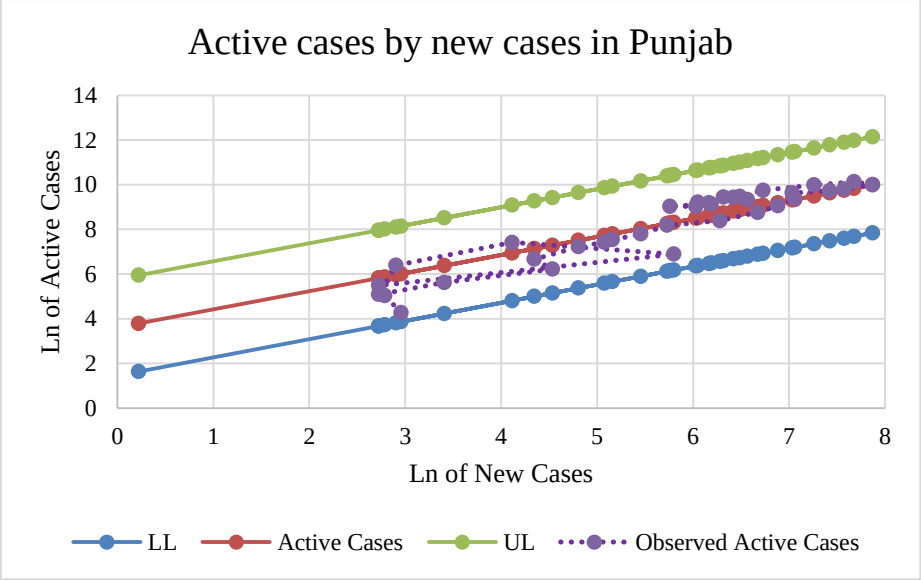


Figure 31: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Punjab

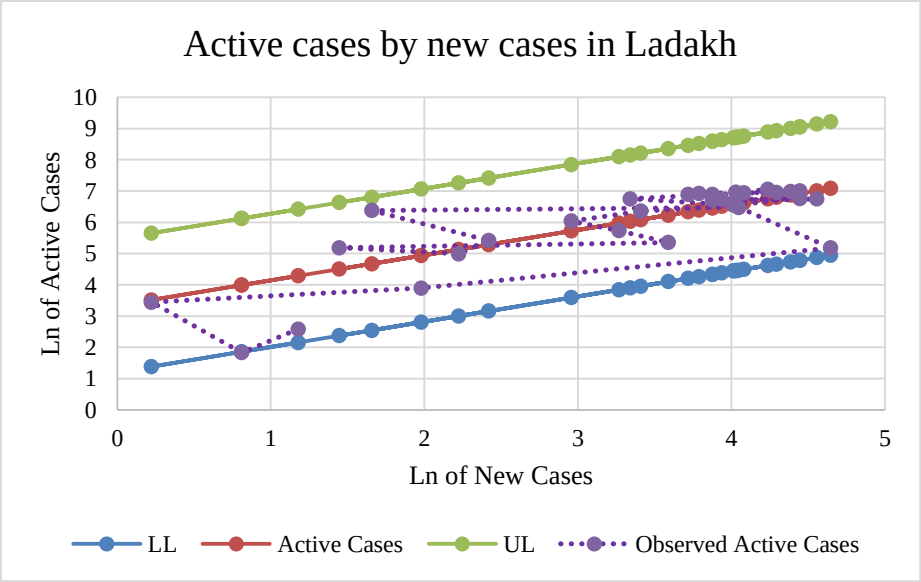


Figure 32: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Ladakh

In the northern region, log of new cases against log of predicted active cases was plotted with the corresponding upper and lower confidence limits for the states Rajasthan, Jammu and Kashmir, Himachal Pradesh, Uttarakhand, Haryana, Delhi (R^2=0.87), Punjab, Chandigarh, and Ladakh and were found to be linear (Figure 25-32). The average duration of days in Rajasthan (R^2=0.93) and Jammu Kashmir (R^2=0.93) is 14 days (CI = 11.801-16.984) (CI = 11.337-16.118), Himachal Pradesh (R^2=0.91) takes around 18 days (CI = 15.930-21.309) while Uttarakhand (R^2=0.90) takes around 20 days (CI = 116.625-23.616) followed by Haryana (R^2=0.89) 31 days (CI = 25.567-38.029) and Chandigarh (R^2=0.78) 32 days (CI = 26.839-37.502). With an average length of 36 days, Punjab (R^2=0.8) has the highest recovery rate among the northern states (CI = 28.370-48.042). Delhi (R^2=0.87) has an average recovery time of 22 days having a confidence interval of 17 days to 30 days. The union territories like Ladakh (R^2=0.75), which has a model fit of 75 per cent (R^2=0.75) and has a duration of 28 days for a patient to be patient (CI = 23.76 and 33.32).

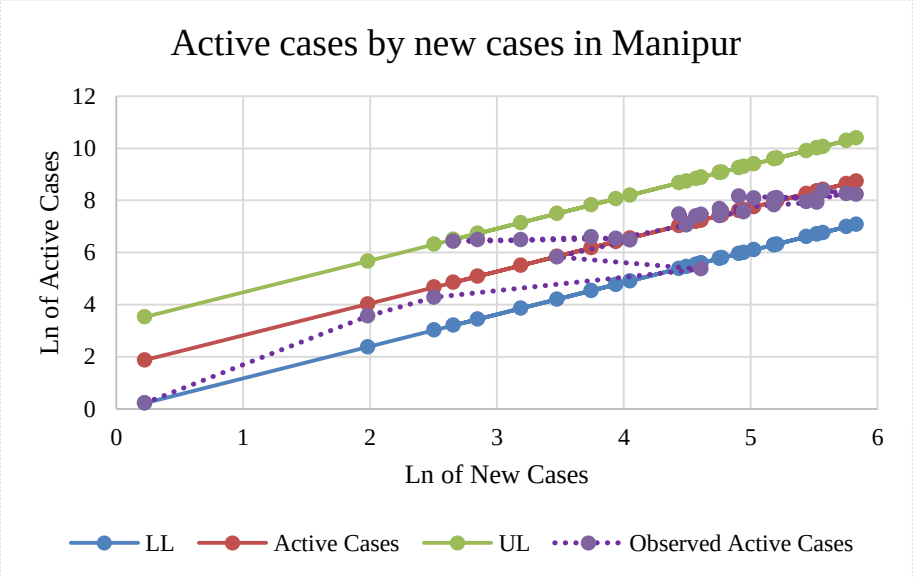


Figure 33: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Manipur

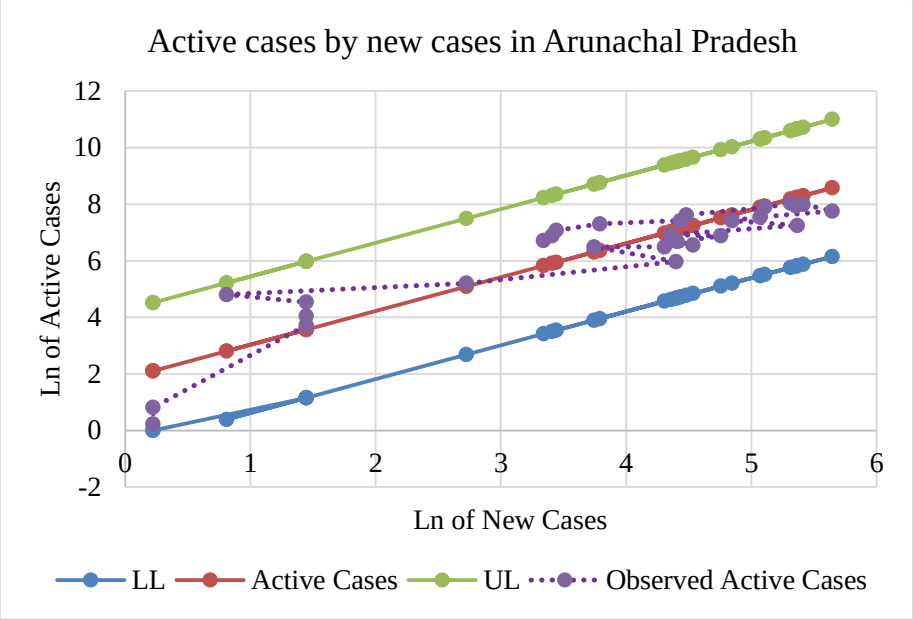


Figure 34: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Arunachal Pradesh

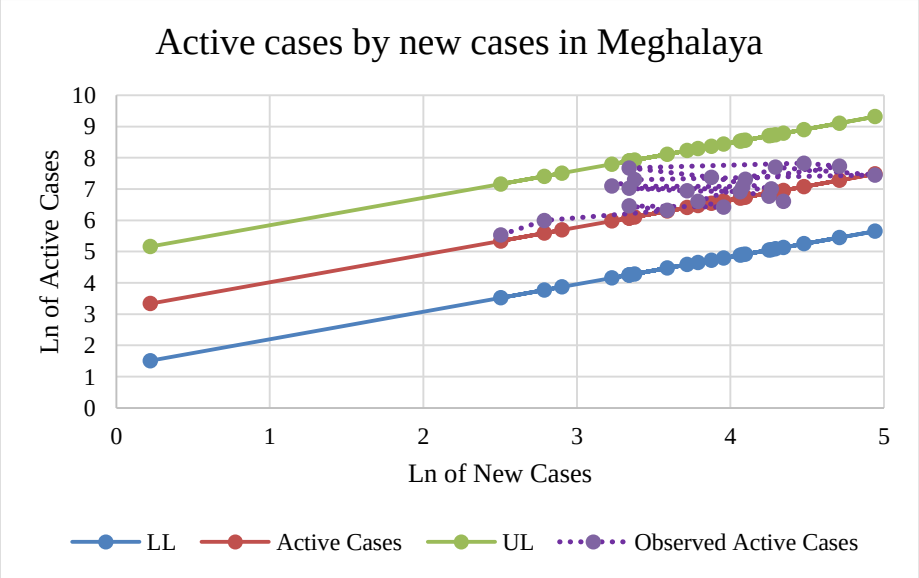


Figure 35: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Meghalaya

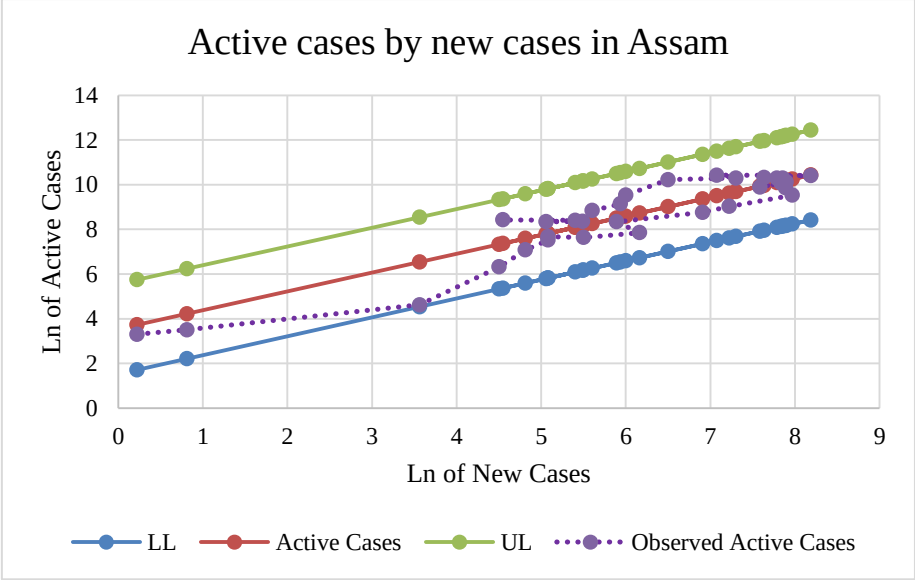


Figure 36: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Assam

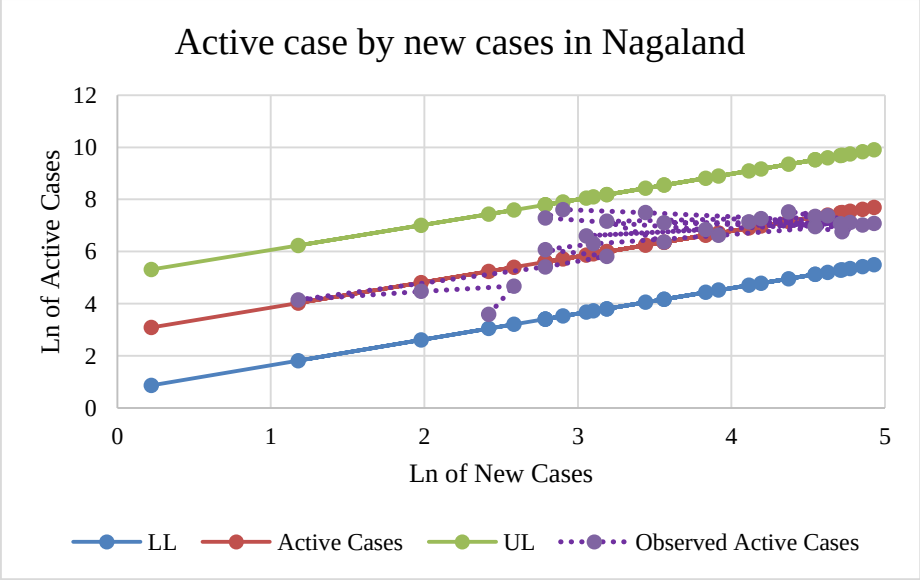


Figure 37: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Nagaland

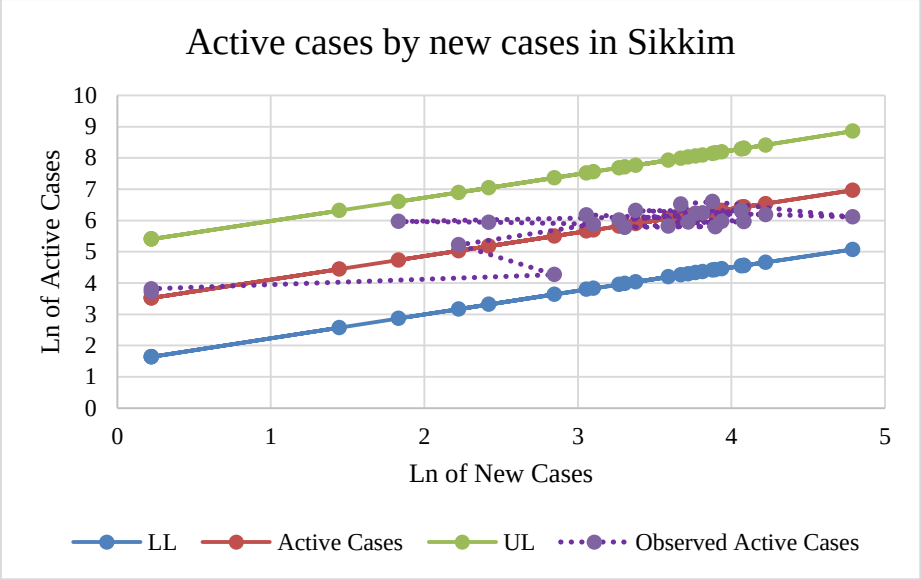


Figure 38: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Sikkim

Log of new cases against log of predicted active cases with the upper and lower confidence limits was plotted for northeast states. Manipur, Arunachal Pradesh, Meghalaya, Assam, Nagaland, Sikkim, and Mizoram (Figure 34-38). High model fit was found in states of Manipur (R^2=0.93), Arunachal Pradesh (R^2=0.86), Meghalaya (R^2=0.85) and Assam (R^2=0.85) with average recovery days of 5, 6, 23, 34 days. Nagaland (R^2=0.67) has a model fit of 67 per cent (R^2=0.67) and has a duration of 18 days (CI = 12.42-24.99), similarly Sikkim(R^2=0.66) with a model fit of 67 percentages (R^2=0.67) and duration of 29 days (CI = 22.84-35.76). Finally, Mizoram (R^2=0.58) has the least model fit with 59 per cent (R^2=0.59) and a duration of 16 days (CI = 12.20-19.78).

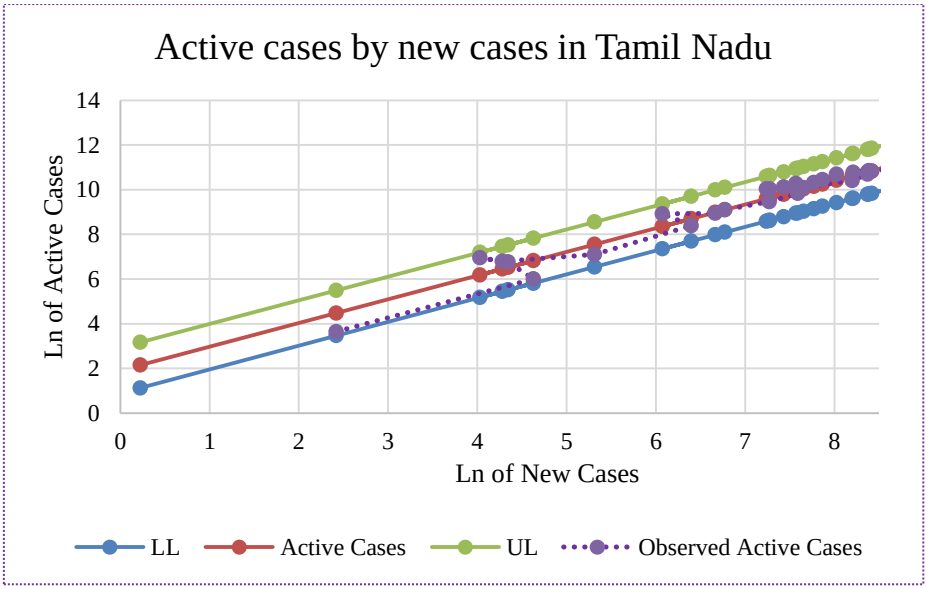


Figure 39: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Tamil Nadu

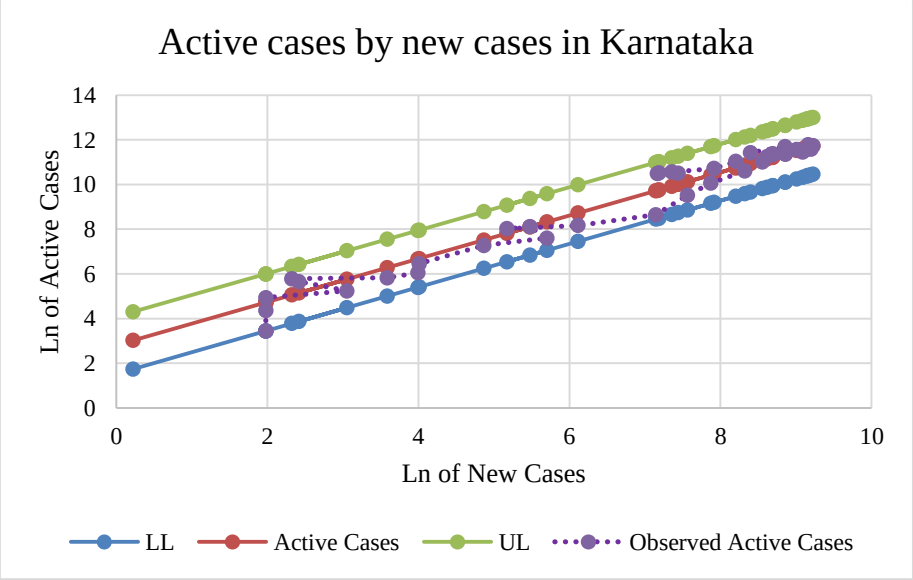


Figure 40: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Karnataka

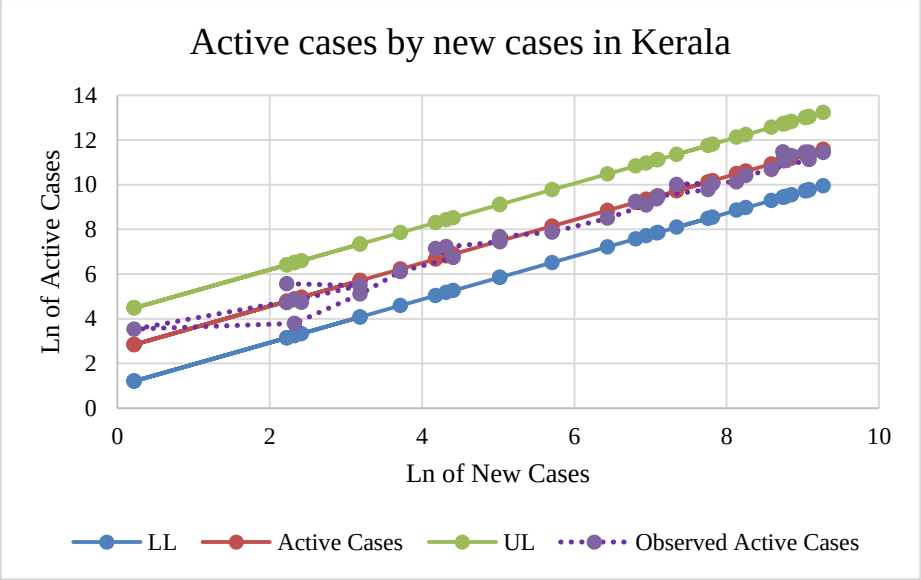


Figure 41: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Kerala

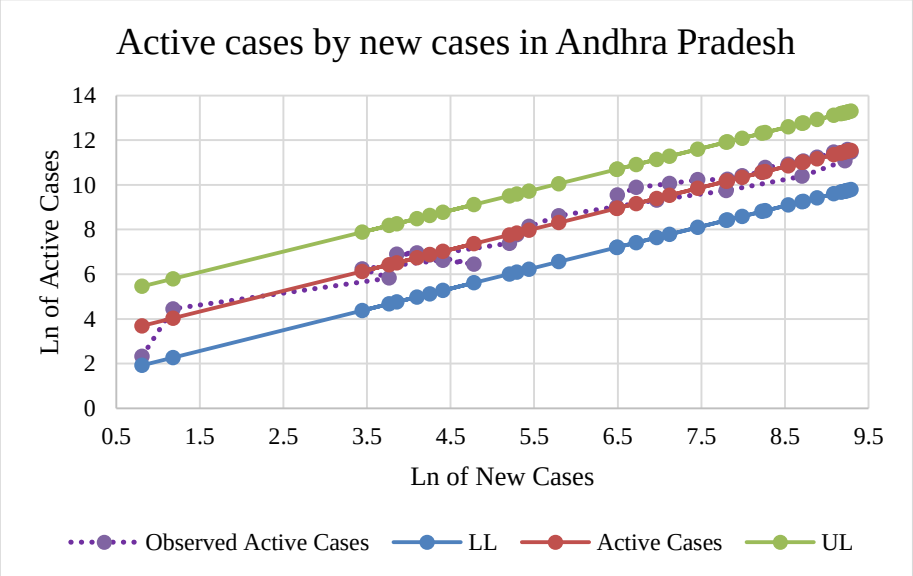


Figure 42: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Andhra Pradesh

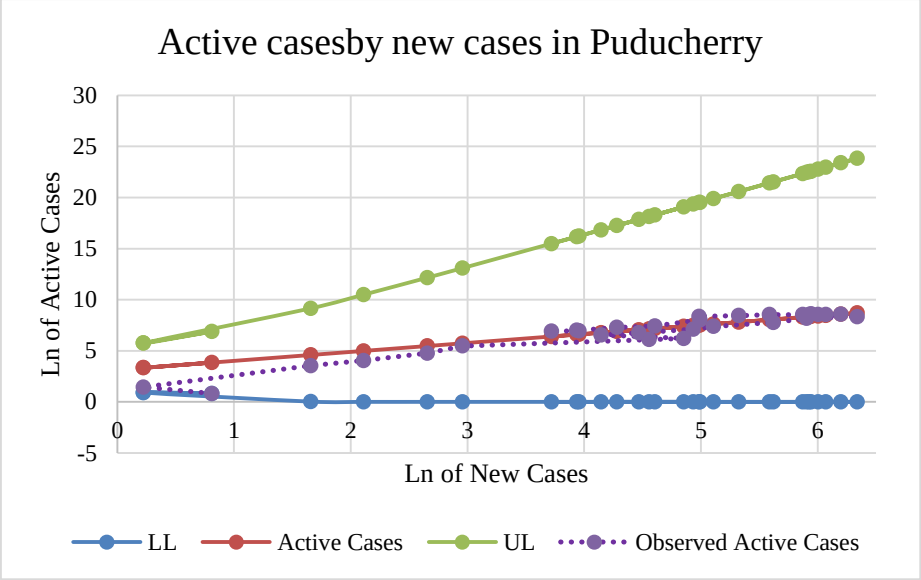


Figure 43: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Puducherry

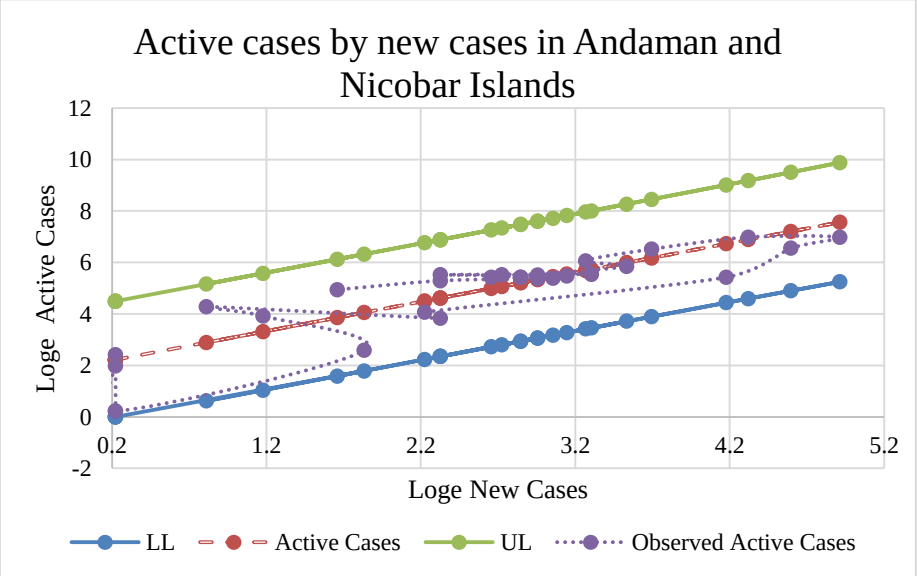


Figure 44: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Andaman and Nicobar Island

For southern states named Tamil Nadu, Karnataka, Kerala, Andhra Pradesh, Puducherry, and Andaman and Nicobar Islands, log of new cases against log of predicted active cases with the upper and lower confidence limits was plotted (Figure 39-44). Andhra Pradesh (R^2=0.89) takes an average of 19 days (CI = 14.610-24.472, 15.642-22.675) to recover, where Andaman and Nicobar Islands (R^2=0.8) and Tamil Nadu (R^2=0.96) take an average of 7 days (CI = 5.993-8.530, 5.710-7.999) to recover, followed by Kerala (R^2=0.94) with 14 days (CI = 11.882-16.170), Karnataka (R^2=0.95) with 17 days (CI = 13.753-19.829), and Puducherry (R^2=0.82) with 23 days (CI = 18.743-28.089).

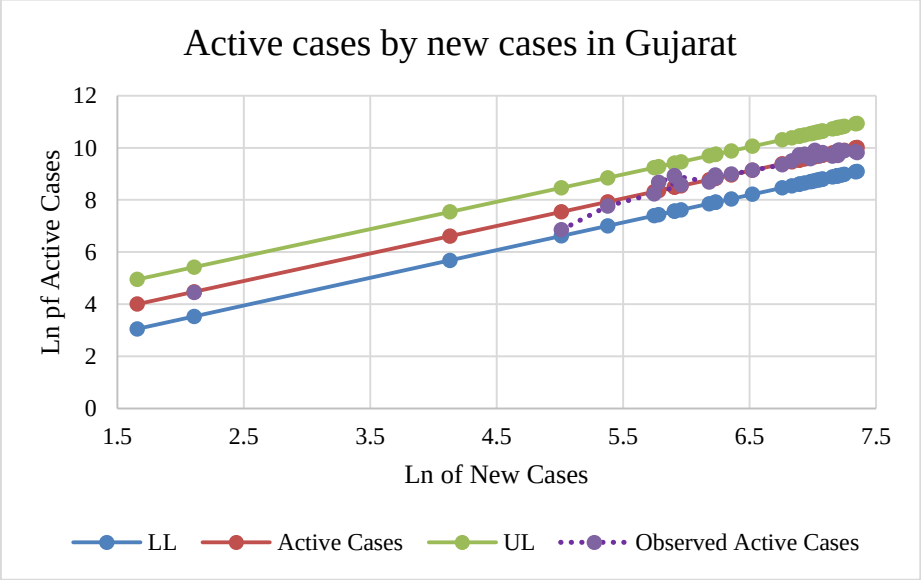


Figure 45: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Gujarat

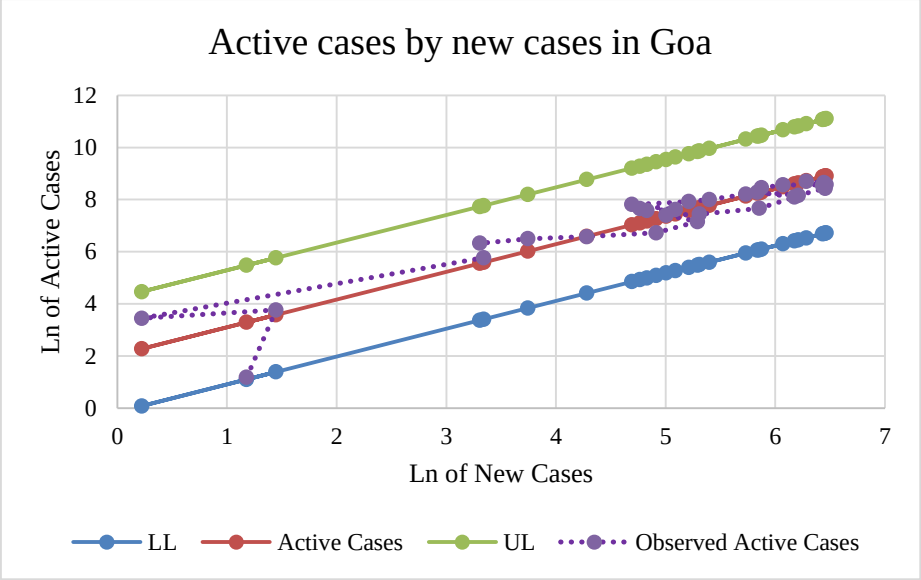


Figure 46: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Goa

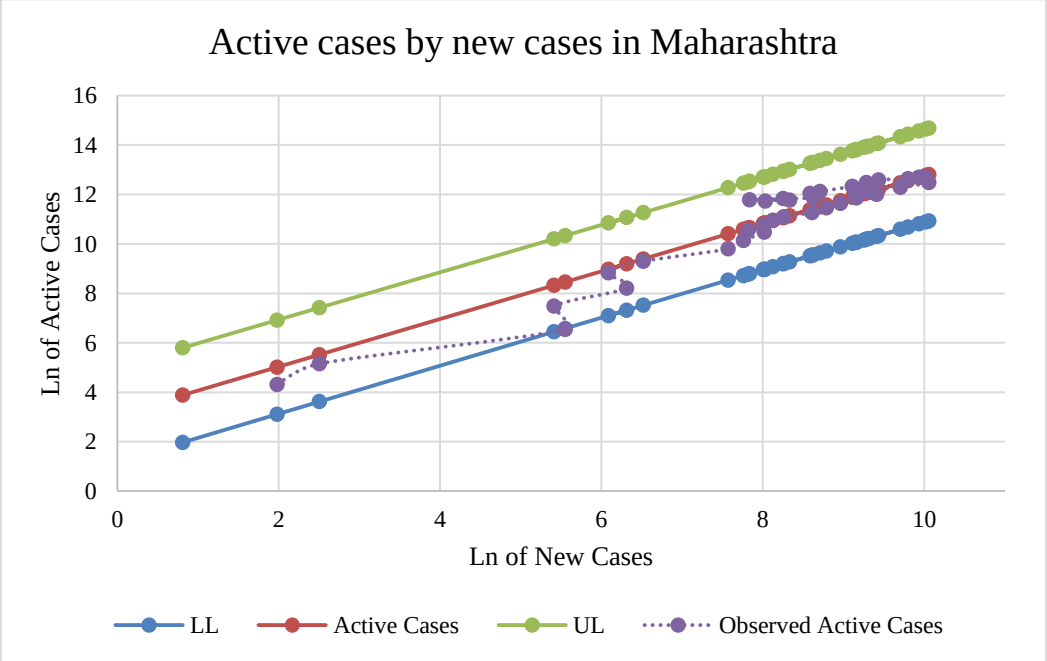


Figure 47: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Maharashtra

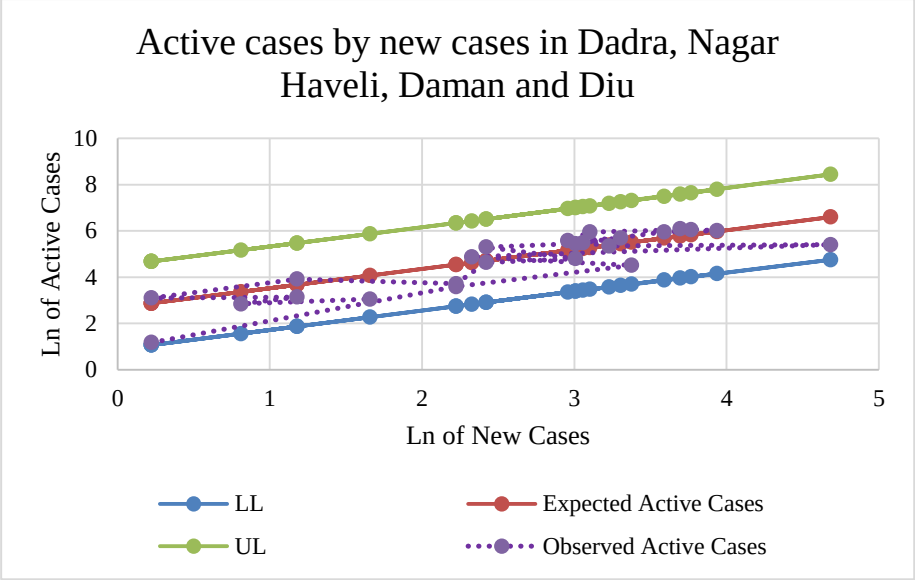


Figure 48: Logarithmic Trend of Active COVID-19 Cases Versus New Cases in Dadra, Nagar Haveli, Daman and Diu

In western states, Gujarat, Goa, Maharashtra Dadra, Nagar Haveli, Daman and Diu, log (New Cases) against log (Predicted Active Cases) along with its upper and lower confidence limits was plotted (Figure 45-48). Dadra and Nagar Haveli and Daman and Diu (R^2=0.73) have a model fit of 73 per cent (R^2=0.73) and has a duration of 15 days with a confidence interval of 12.44 days and 17.35 days wherein Goa (R^2=0.88) and Gujarat (R^2=0.92) the duration is almost equal, almost 8 to 10 days (CI = 7.450-12.098, 6.167-9.524). Maharashtra (R^2=0.86) has and average recovery time of 22 days with a confidence interval of 15 days to 32 days (Table 21).

Table 21: Average duration (days) of COVID-19 patients in Indian states

Region	State	R^2	F	B	b_1	AD	LL	UL
Central	Chhattisgarh	0.941	4184.1	2.51	0.969	12.3	10.4	14.5
	Uttar Pradesh	0.883	2101.0	3.22	0.896	25.0	19.3	32.4
	Madhya Pradesh	0.761	835.5	4.22	0.726	68.3	50.3	92.8
East	Odisha	0.956	5823.2	2.56	0.940	12.9	11.1	15.0
	West Bengal	0.948	4789.7	2.44	0.980	11.4	9.4	13.8
	Bihar	0.923	3143.0	2.80	0.914	16.5	13.6	20.1
	Jharkhand	0.871	1695.1	3.20	0.867	24.6	19.8	30.6
North	Rajasthan	0.937	4179.2	2.65	0.952	14.1	11.8	16.9
	J&K	0.932	3737.5	2.60	1.007	13.5	11.3	16.1
	Himachal Pradesh	0.915	2802.9	2.91	0.898	18.4	15.9	21.3
	Uttarakhand	0.901	2441.0	2.98	0.899	19.8	16.6	23.6
	Haryana	0.891	2274.5	3.44	0.806	31.1	25.5	38.0
	Delhi	0.870	1884.8	3.12	0.892	22.7	17.1	30.0
	Punjab	0.802	1112.4	3.60	0.811	36.9	28.3	48.0
	Chandigarh	0.789	986.2	3.45	0.740	31.7	26.8	37.5
	Ladakh	0.750	829.7	3.33	0.806	28.1	23.7	33.3
Northeast	Manipur	0.936	3776.9	1.59	1.224	4.9	4.2	5.8
	Arunachal Pradesh	0.863	1568.9	1.83	1.193	6.2	5.0	7.7
	Meghalaya	0.856	1414.2	3.14	0.878	23.1	19.7	27.0
	Assam	0.850	1423.1	3.54	0.842	34.4	26.6	44.5
	Nagaland	0.671	417.5	2.86	0.980	17.6	12.4	24.9
	Sikkim	0.669	401.0	3.35	0.754	28.5	22.8	35.7
	Mizoram	0.588	368.8	2.74	0.955	15.5	12.2	19.7
South	Tamil Nadu	0.967	8117.1	1.91	1.061	6.7	5.7	7.9
	Karnataka	0.950	5227.6	2.80	0.968	16.5	13.7	19.8
	Kerala	0.946	5514.6	2.62	0.966	13.8	11.8	16.1
	Andhra Pradesh	0.898	2381.9	2.94	0.926	18.9	14.6	24.4
	Telangana	0.840	2767.2	2.93	0.935	18.8	15.6	22.6
	Puducherry	0.826	1257.3	3.13	0.878	22.9	18.7	28.0
	A&N	0.804	1052.7	1.96	1.139	7.1	5.9	8.5
West	Gujarat	0.920	3043.2	2.25	1.056	9.4	7.4	12.0
	Goa	0.884	1949.5	2.03	1.064	7.6	6.1	9.5
	Maharashtra	0.864	1746.5	3.10	0.965	22.2	15.3	32.1
	DNH&D	0.732	593.0	2.68	0.835	14.6	12.4	17.3

Note: p value <0.05 for all states, B= unstandardized beta coefficient, b1= slop of the regression line, AD- Average Duration (days), LL- Lower Limit, UL- Upper Limit, J&K- Jammu and Kashmir, A&N- Andaman and Nicobar, DNH&D- Dadra, Nagar Haveli, Daman and Diu

3.2 Model

3.2.1. Basic Model

Table 22: COVID-19 prediction in Exponential Model, Logistic Model, Gompertz Model, and Bertalanffy model.

Model	Parameter	Value Of the Parameter
Exponential model	I_0	0.399
	r	0.128
Logistic model	I_0	13642.053
	r	0.094
	k	124935.667
Gompertz model	I_0	20.945
	r	0.021
	k	789143.13
Bertalanffy model	I_0	12.333
	r	0.014
	k	1609200.51

Note: Where $\boldsymbol{I_0}$ is the initial number of cases, $\boldsymbol{k}$ is the carrying capacity of an epidemic, $\boldsymbol{r}$ is the growth rate

Table 23: Coefficient of determination ($R^{2)}$ based on fitted models.

Growth Models	Fitted Growth Models	R^2 Values
Exponential growth model	$I(t) = 0.399 * e^{0.128*t}$	0.160
Logistic growth model	$I(t) = \frac{0.094}{1 + 13642.053 * e^{-124935.667*t}}$	0.997
Gompertz growth model	$I(t) = 789143.136 * e^{-20.945*e^{-0.021t}}$	0.998
Bertalanffy growth model	$I(t) = 1609200.51(1 - e^{-0.0146t})^{12.3339}$	0.999

Referring Table 22 and Table 23, for exponential model (Figure 49), the expected number of infected cases at time t = 0, I_0 is 0.399 and the value of the growth factor r is 0.1280. By substituting these values on the exponential model, we get,

$$I(t) = 0.399 * e^{0.128*t}$$

This formula was used to estimate the predictions and also experienced correlation score 0.160 between predicted and actual scores. This correlation score clearly highlights that exponential model is inefficient to estimate the cumulative spread of COVID-19 pandemic in future days. Hence moving on to higher growth models.

The prediction methods of Logistic model, Gompertz model and Bertalanffy model are similar, but the mathematical models are different. For logistic model (Figure 50), 1,24,935.667 is the predicted maximum of the confirmed cases and the value of the growth factor r is 0.094. The expected number of infected cases at time t = 0, I_0 is 0.094. By substituting these values on the Logistic model we get,

$$\mathrm{I(t)} = \frac{0.094}{1+13642.053*\mathrm{e}^{-124935.667*\mathrm{t}}}.$$

This formula was used to estimate the predictions and also experienced correlation score 0.997 between predicted and actual scores. It indicates that logistic model is 99.76% efficient in making predictions.

For Gompertz model (Figure 51), 7,89,143.1360 is the predicted maximum of the confirmed cases and the value of the growth factor r is 0.021. The expected number of infected cases at time t = 0, I_0 is 20.945. By substituting these values on the logistic model we get,

$$I(t) = 789143.136 * e^{-20.945 * e^{-0.021t}}$$

This formula was used to estimate the predictions and also experienced correlation score 0.998 between predicted and actual scores. It indicates that logistic model is 99.89% efficient in making predictions.

For Bertalanffy model (Figure 52), 16,09,200.51 is the predicted maximum of the confirmed cases and the value of the growth factor r is 0.014. The expected number of infected cases at time t = 0, I_0 is 12.333. By substituting these values on the logistic model we get,

$$I(t) = 1609200.51(1 - e^{-0.014t})^{12.333}.$$

This formula was used to estimate the predictions and also experienced correlation score 0.999 between predicted and actual scores. It indicates that logistic model is 99.9% efficient in making predictions. This correlation score clearly highlights that Bertalanffy model is most efficient in estimating the cumulative spread of COVID-19 pandemic in future days.

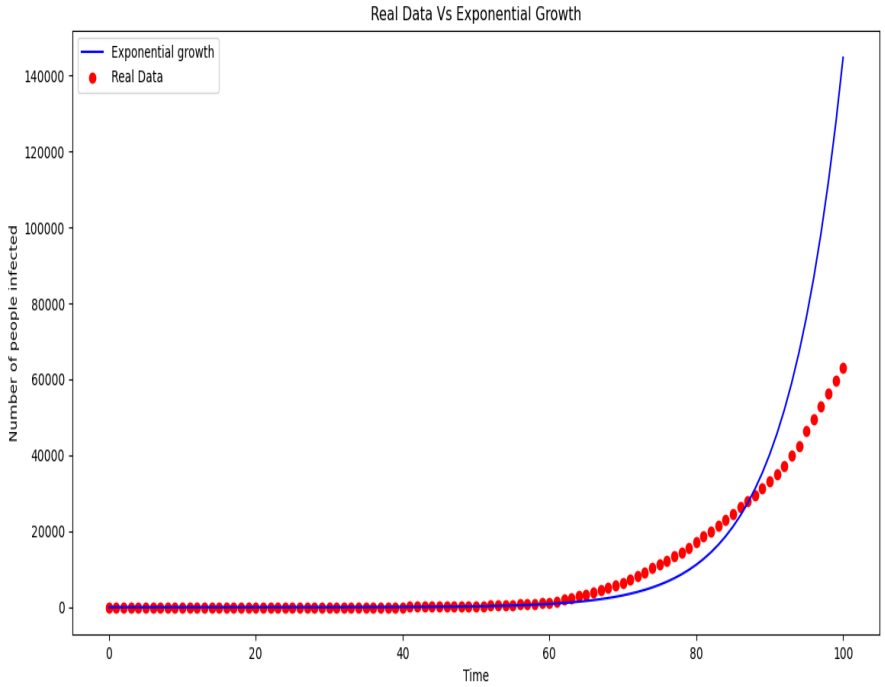


Figure 49: Exponential growth curve of COVID-19 in India

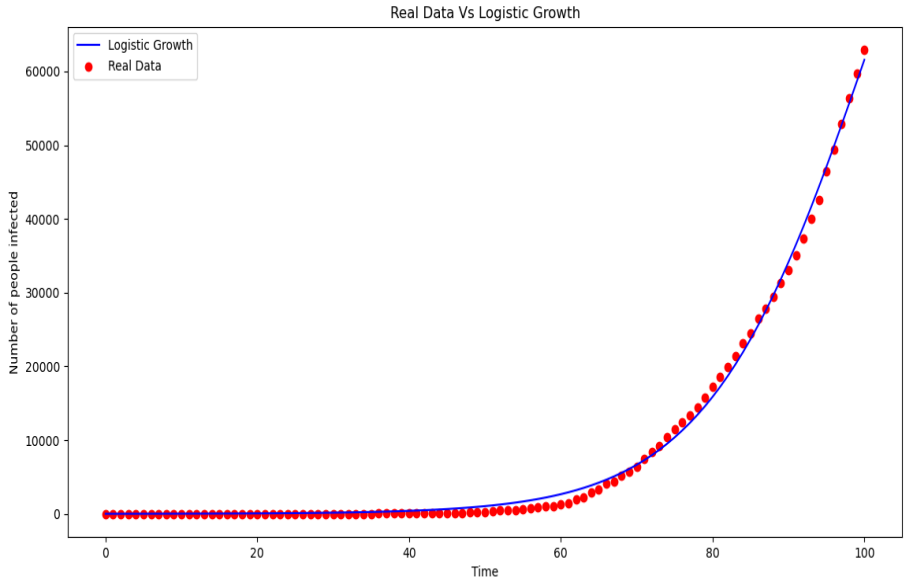


Figure 50: Logistic growth curve of COVID-19 in India

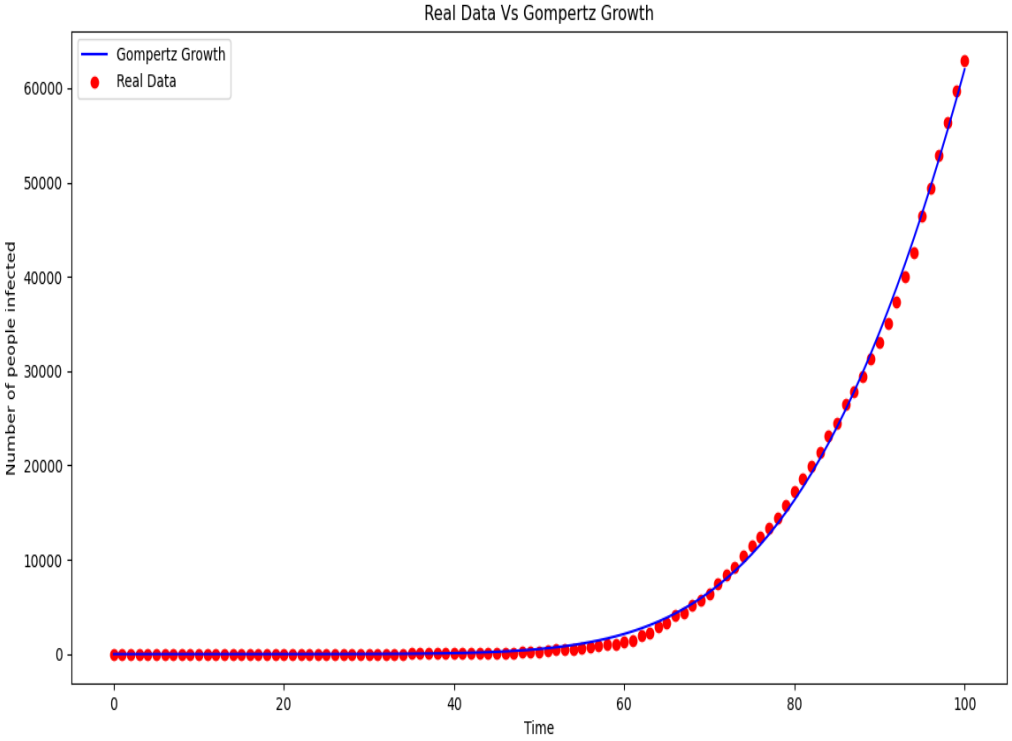


Figure 51: Gompertz growth curve of COVID-19 in India.

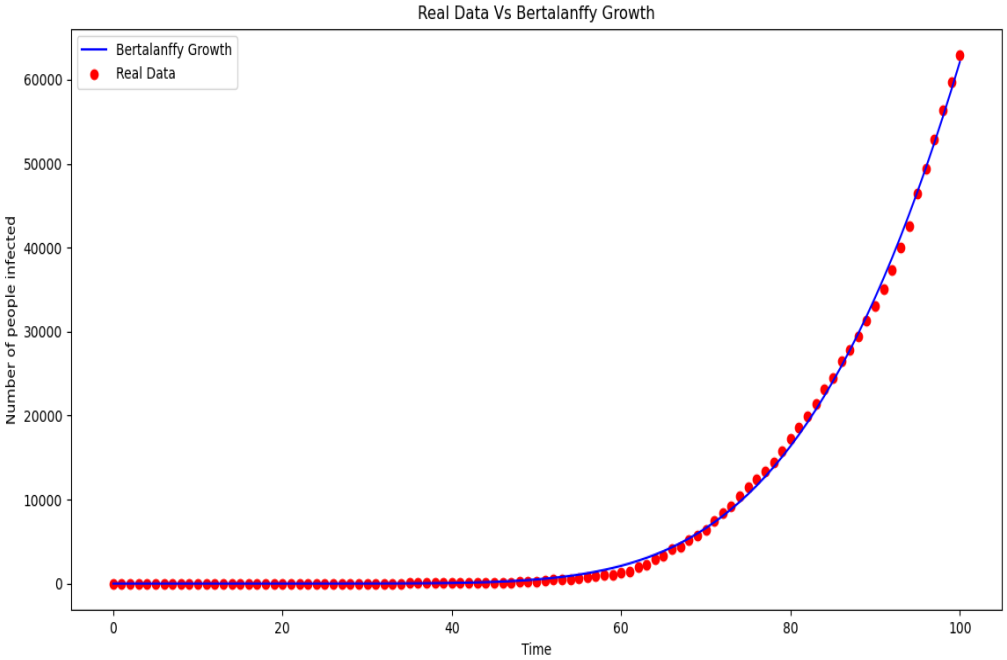


Figure 52: Bertalanffy growth curve of COVID-19 in India.

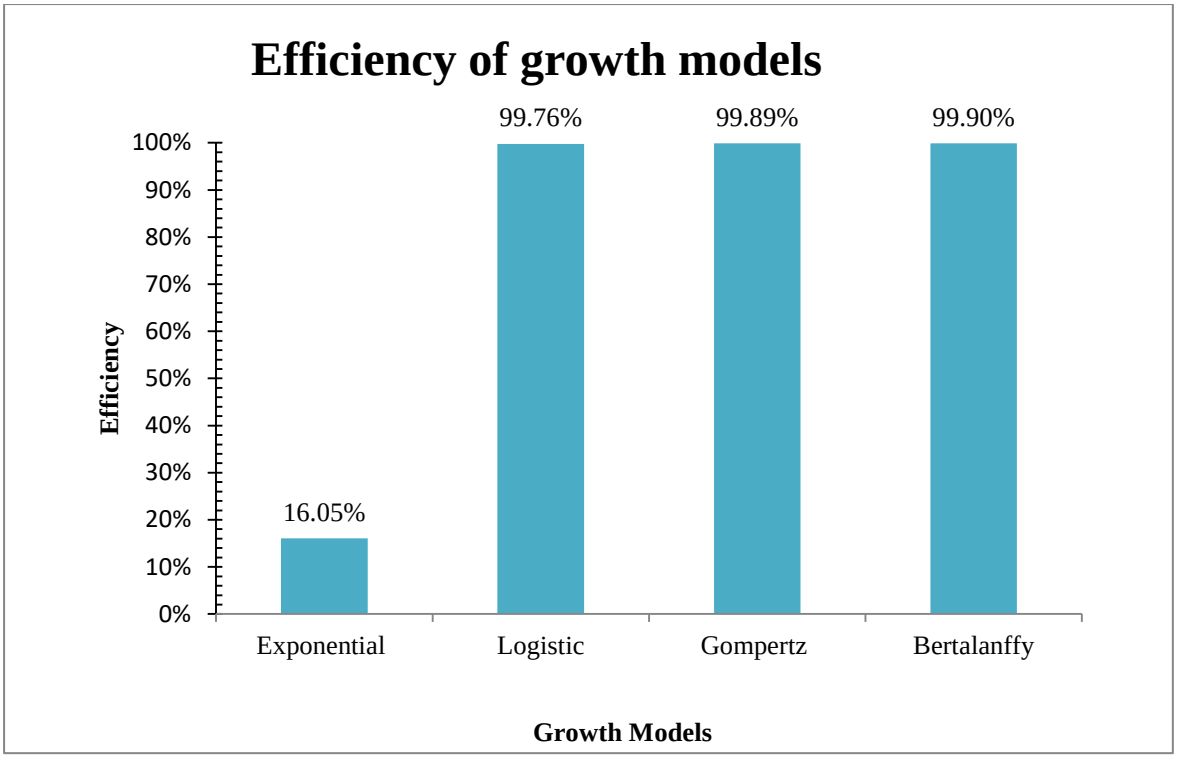


Figure 53: Efficiency of growth models.

According to the value of R^2, the best model fitted is Bertalanffy growth model (R^2=0.999) (Figure 53).

Table 24: Observed and Predicted values as per Bertalanffy growth model from 11 May 2020 to 20 June 2020.

Date	Real data	Predicted data	Difference
11-05-2020	67152	65113	2039
12-05-2020	70756	68553	2203
13-05-2020	74281	72114	2167
14-05-2020	78003	75797	2206
15-05-2020	81970	79602	2368
16-05-2020	85940	83529	2411
17-05-2020	90927	87580	3347
18-05-2020	96169	91754	4415
19-05-2020	101139	96053	5086

20-05-2020	106750	100075	6675
21-05-2020	112359	105022	7337
22-05-2020	118447	108692	9755
23-05-2020	125101	112486	12615
24-05-2020	131868	116404	15464
25-05-2020	138845	121444	17401
26-05-2020	145380	126607	18773
27-05-2020	151767	131891	19876
28-05-2020	158333	137297	21036
29-05-2020	165799	142822	22977
30-05-2020	173763	148467	25296
31-05-2020	182143	154229	27914
01-06-2020	190535	160109	30426
02-06-2020	198706	167104	31602
03-06-2020	207615	174213	33402
04-06-2020	216919	181434	35485
05-06-2020	226770	187768	39002
06-06-2020	236657	194210	42447
07-06-2020	246628	200761	45867
08-06-2020	256611	207418	49193
09-06-2020	266598	214179	52419
10-06-2020	276583	221042	55541
11-06-2020	286579	228006	58573
12-06-2020	297535	235068	62467
13-06-2020	308993	242227	66766
14-06-2020	320922	249479	71443
15-06-2020	332424	256824	75600
16-06-2020	343091	264258	78833
17-06-2020	354065	271780	82285
18-06-2020	366946	279387	87559
19-06-2020	380532	287076	93456
20-06-2020	395048	294846	100202

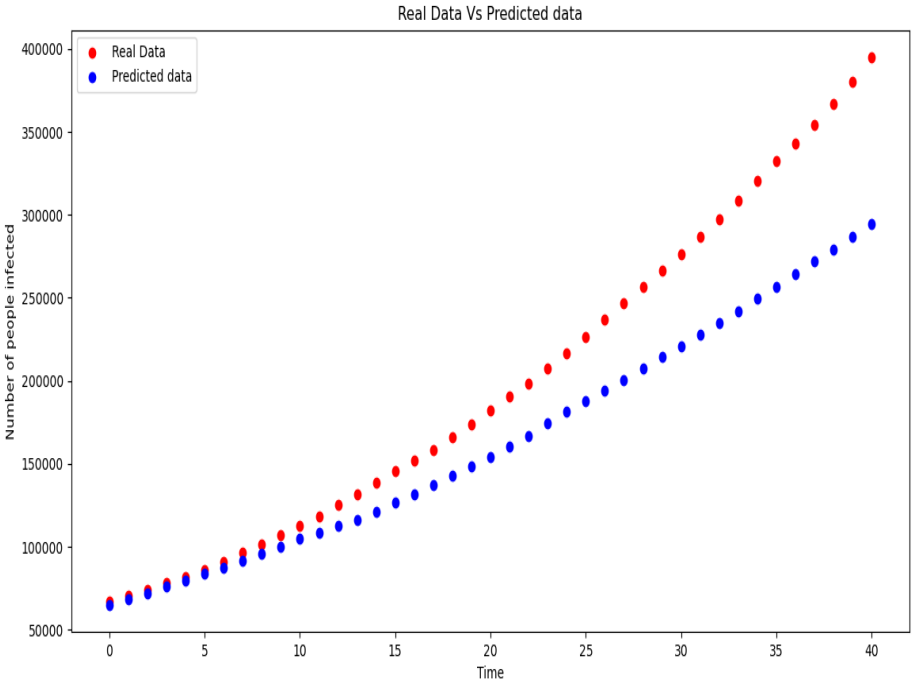


Figure 54: Graph showing deflection between observed data vs. predicted data.

Table 24 is the comparison between observed and predicted data according to the Bertalanffy growth model for the period from May 11, 2020, to June 20, 2020. The table includes dates, real data, predicted data, and the difference between the two, highlighting how well the model aligns with the actual COVID-19 case numbers during this timeframe. From its graphical representation (Figure 54) it is clear that after a certain period of time there is a clear deviation from the predicted number of cases. The deviation is resulted from the violation of growth model assumption that was made on the population which includes the phase-wise removal of lockdown, arrival of expats and partial lifting of restrictions. As on 20 June 2020 the predicted number of cases and actual number of cases differs in 100202. This huge deviation due to the ill-timed relaxation of lockdown. This will lead to further research question such as whether the lockdown was successful in reducing the impact of COVID-19 in India.

3.2.2. Prediction Model

3.2.2.1 Time-Response and Viral Shedding of COVID-19

Shown below the obtained values of MR, RR, Crude CFR, and the Yoshikura estimate of CFR. The data required was collected from March 12, 2020 to July 3, 2020. Kerala has the least CFR of 0.52% indicating the least severity and Gujarat has the highest CFR of 5.49% representing the acme of the epidemic. The states with comparatively low CFR other than Kerala are Himachal Pradesh (0.87%) and Andhra Pradesh (1.22%). And Maharashtra (4.34%) is the state with comparatively high CFR other than Gujarat. Along with CFR two other measures namely Mortality Rate and Recovery Rate to comprehend and analyze the impact of COVID-19.

Table 25: The values obtained for MR, RR, Crude CFR, Yoshikura CFR from March 12, 2020 to July 3, 2020.

STATE	MR	RR	CRUDE CFR	YOSHIKURA CFR
Kerala	0.074	59.051	0.00522	0.00523
Tamil Nadu	1.919	56.837	0.01348	0.00753
Jammu & Kashmir	0.970	63.361	0.01486	0.00921
Punjab	0.569	71.885	0.02655	0.02594
Andhra Pradesh	0.243	45.457	0.01227	0.01337
Karnataka	0.482	44.682	0.01496	0.01243
Himachal Pradesh	0.131	61.179	0.00872	0.00746
Telangana	0.804	49.963	0.01387	0.0171
Maharashtra	7.452	54.255	0.04344	0.03676
Gujarat	3.153	71.934	0.05497	0.05393

Kerala was observed with the lowest mortality rate of 0.07483 and highest recovery rate of 59.05%. At the same time Maharashtra was found to be one of the severely affected states with the most elevated death pace of 7.45276 (approx.) (Table 25). The

graphical comparison of Crude CFR and Yoshikura CFR from March 12, 2020 to July 3, 2020 (Figure 55) along with the Mortality rate and Recovery rate of states from March 12, 2020 to July 3, 2020 are presented (Figure 56).

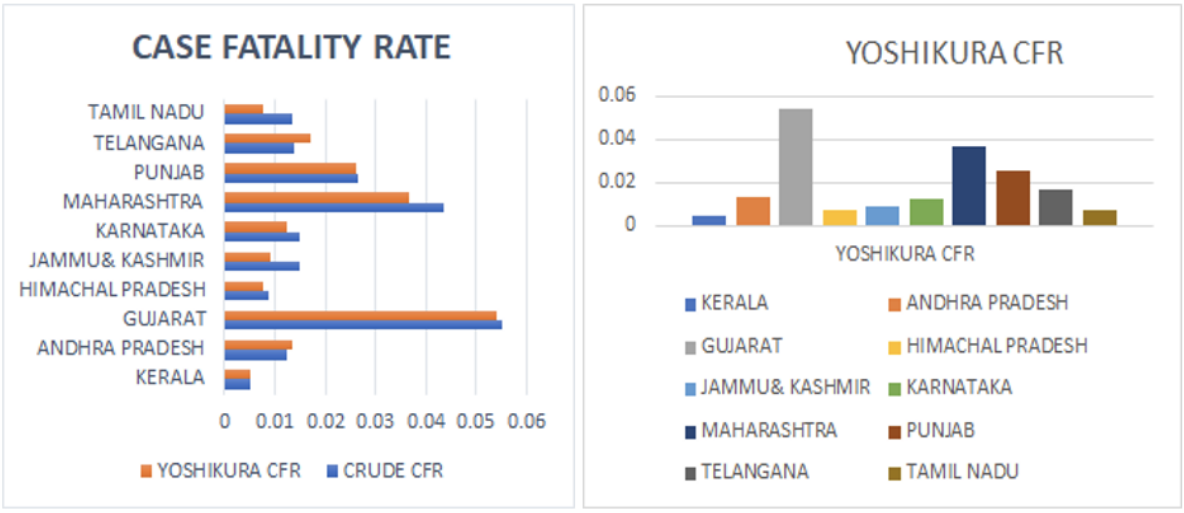


Figure 55: Comparison of Crude CFR and Yoshikura CFR from March 12, 2020 to July 3, 2020.

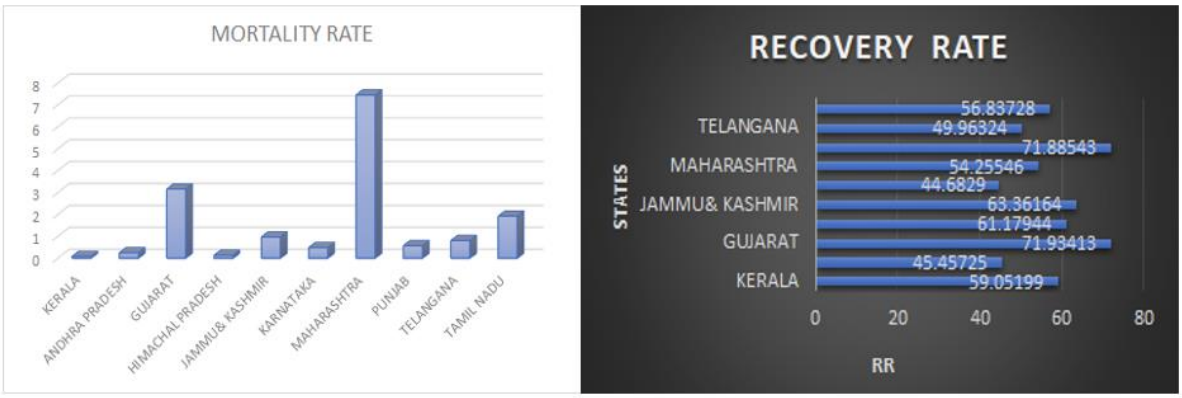


Fig 56: Mortality rate and Recovery rate of states from March 12, 2020 to July 3, 2020

For evaluating the statistical distribution that best fits the data, 73 daily data from 12 Mar 2020 of India. All the variables are summarized below using descriptive statistics. Table 26 reflects the descriptive statistics with a skewness and kurtosis of

4.97344 and 41.16669 respectively indicating the deviation from normality to right skewed distribution. Fig 57 depicts the histogram and CDF plot indicating the right skewed distribution graphically. Fig 58 depicts the Skewness-kurtosis plot for a continuous variable. Fig 59 portrays the goodness of fit plots of Weibull, Gamma, and Lognormal distributions. These plots helped us to determine the best fitted distribution.

Table 26: COVID-19 Mortality in India.

Data	Min	Max	Median	Mean	Estimated S.D	Estimated skewness	Estimated kurtosis
Number of Deaths	0	2004	103.5	163.78	226.18	4.97	41.16

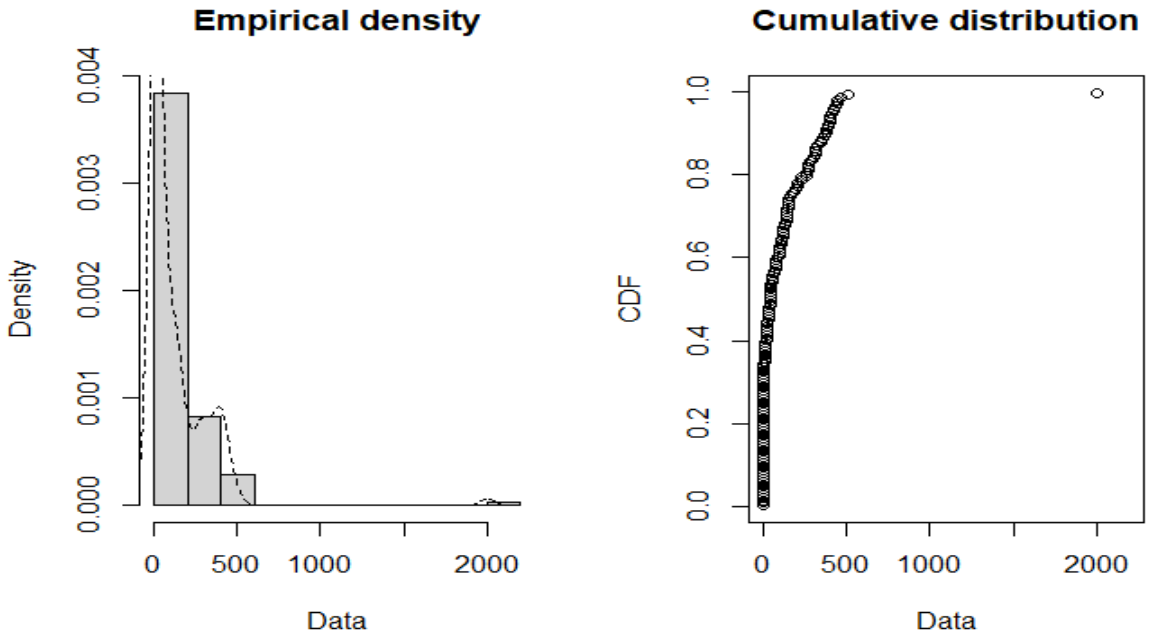


Figure 57: Histogram and CDF plots of an empirical distribution for a continuous variable (fitted to the data of COVID 19 mortality in India)

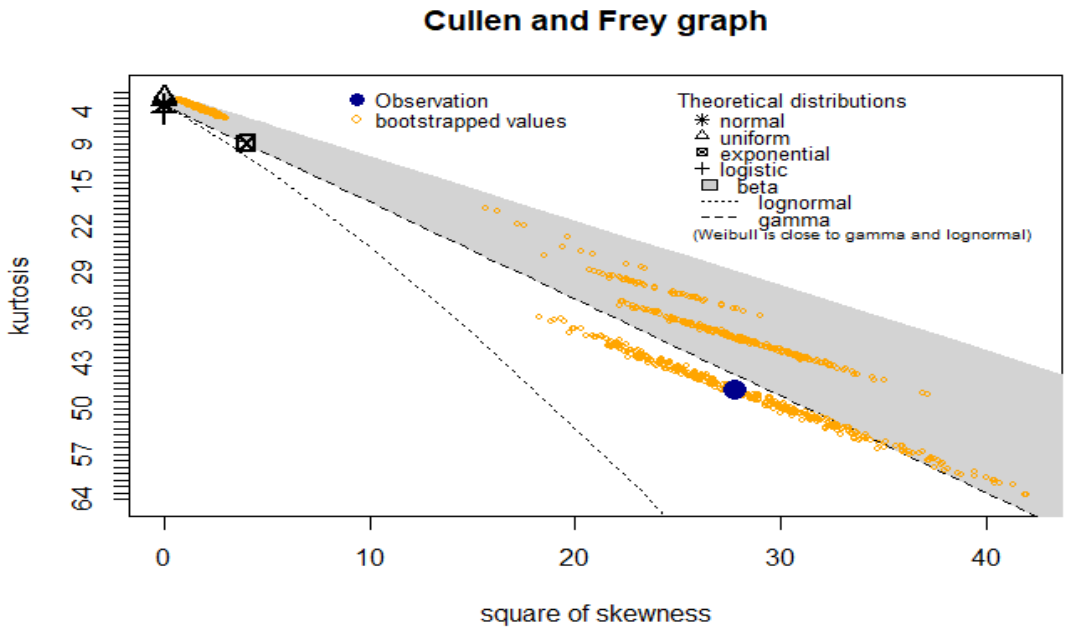


Figure 58: Skewness-kurtosis plot for a continuous variable (fitted to the data of COVID 19 mortality in India)

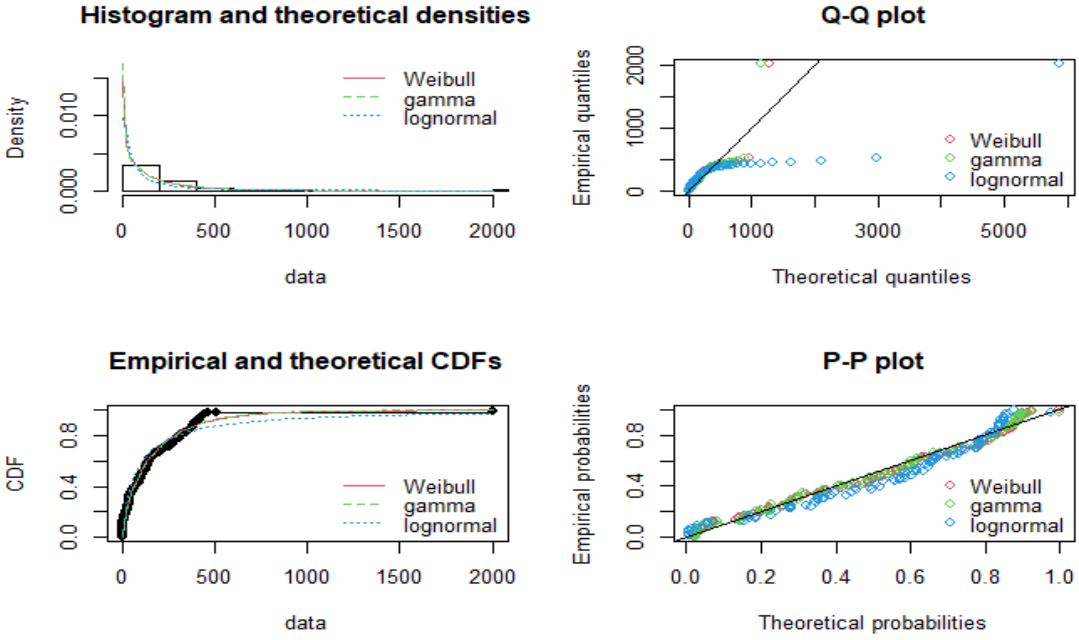


Figure 59: Four goodness-of-fit plots for various distributions fitted to continuous data (Weibull, Gamma and Lognormal distributions fitted to the data of COVID 19 mortality in India).

Table 27 demonstrates the estimate and standard error of the unknown parameters of Weibull, Gamma and Lognormal distributions along with the Log Likelihood, AIC and BIC values using the MLE method. Hence secured the gamma distribution owns the maximum Log likelihood value.

Table 27: COVID-19 Mortality Distributions in India Using MLE Method

Distributions	**Parameters**			**Log likelihood**	**AIC**	**BIC**
Weibull		Estimate	Std Error	-659.17	1322.35	1327.72
	Shape	0.79	0.06			
	Scale	152.43	19.43			
Gamma	Shape	0.68	0.07	-658.69	1322.35	1327.72
	Rate	0.003	0.001			
Log normal	Mean log	4.27	0.16	-671.59	1347.19	1352.55
	Sdlog	1.69	0.11			

Table 28 reflects the goodness of fit using Kolmogorov-Smirnov statistic. Hence obtain the Gamma distribution is the best fitted distribution as it satisfies the criteria of having lower AIC value of 1321.395 when compared to others.

Therefore Table obtained shape parameter α = 0.68809 and rate parameter λ = 0.00398

Therefore, the fitted model of Gamma distribution is obtained as follows:

$$f(x,\alpha,\lambda)=0.016936x^{-0.31191}e^{-0.00398\,x} \qquad \text{with } x\in R^{+}$$

Table 28: Goodness of fit comparison via K-S Statistic and evaluation through AIC.

Goodness-of-fit statistics			
	Weibull	Gamma	Lognormal
Kolmogorov-Smirnov statistic	0.077	0.070	0.130
Goodness-of-fit criteria			
	Weibull	Gamma	Lognormal
AIC	1322.359	1321.395	1347.191

For forecasting the future COVID-19 death, the daily death of India has been collected from March-10 to May-21. The time series data is analysed using the ARIMA model. Here Auto-ARIMA which builds high-performance models with least AIC value. Here ARIMA (0,1,1) with AIC value 577.45 is obtained. The model is then evaluated by residual analysis. Figure 60 represents the time series plot of COVID-19 deaths in India.

The fitted model of ARIMA (0,1,1) is given as:

$$Z_t = Z_{t-1} + a_t + 0.0934\, a_{t-1}$$

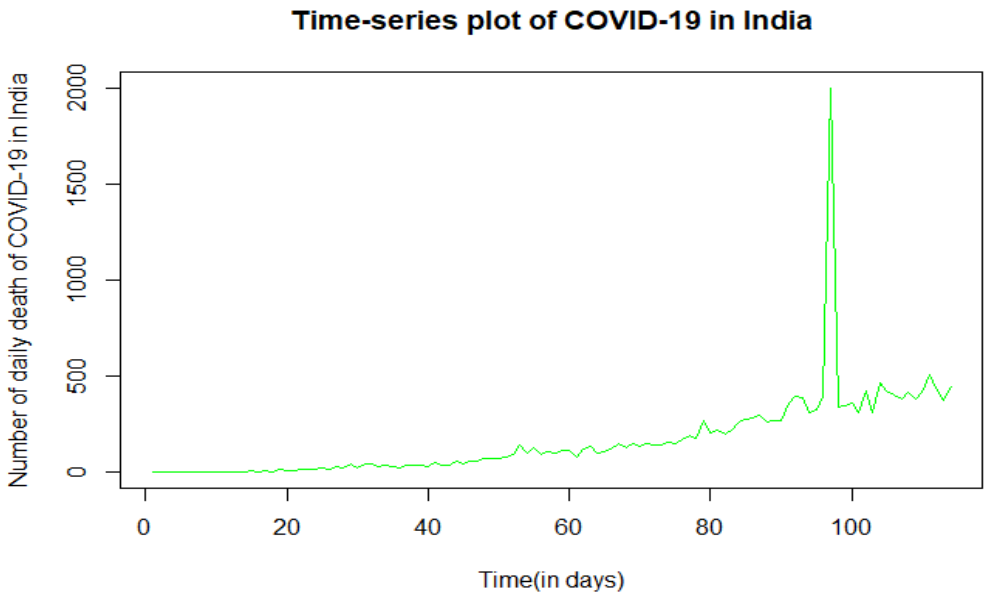


Figure 60: Time series plot of COVID-19 Deaths in India.

Checking for Stationarity

H_0: The data is not stationary

H_1: The data is stationary

Stationarity is checked using Augmented Dickey Fuller Test. The null hypothesis that data is not stationary is tested against alternative hypothesis that the data is stationary. The level of significance (LOF) is taken as 0.05. If p value exceeds LOF, accept the null hypothesis and if the p value lies within the LOF, else reject the null hypothesis that the data is not stationary. If data is not found to be stationary differencing is used to make the mean of the time series which will make the time series data stationary by eliminating trend and seasonality. If single differencing does not attain stationarity, it was repeated until the data becomes stationary

Augmented Dickey-Fuller Test

Dickey-Fuller = -3.5828, Lag order = 4, p-value = 0.03791

Here, p-value>0.05. So, accept the null hypothesis that the data is not stationary. To resolve this difficulty, the method of differencing is used.

Again, checking the stationary of this data using ADF test

Augmented Dickey-Fuller Test

Dickey-Fuller = -7.8113, Lag order = 4, p-value = 0.01

Here p-value<0.05. So, reject the null hypothesis. Thus, the data is Stationary. Next, plot ACF and PACF to find the correlation (Figure 61).

ACF and PACF Plot

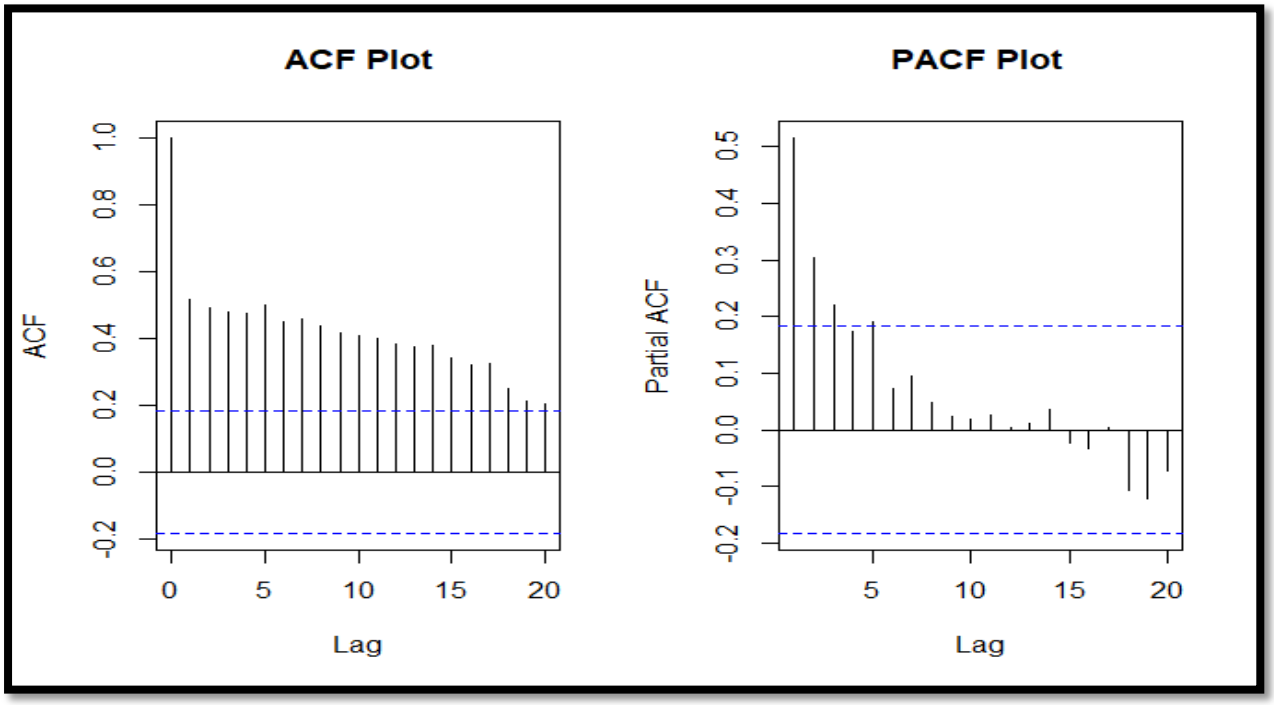


Figure 61: ACF and PACF plot of COVID-19 Deaths in India.

Model estimation using auto. ARIMA

Model estimation using auto arima gives ARIMA (0,1,1) with p =1, d =1 and q =0 with least AIC value of 1480.55

sigma^2 estimated as 27285: log likelihood=-737.27

AIC=1480.55 AIC=1480.77 BIC=1488.73

Checking the model adequacy

a) Stationarity

Here, ADF test is used.

H_0 : The residuals are not stationary

H_1 : The residuals are stationary

Augmented Dickey-Fuller Test

Dickey-Fuller = -4.613, Lag order = 4, p-value = 0.01

Here p-value<0.05. So, residuals are stationary.

b) Checking the Lack of Fit Using Box-Ljung Test

To check the adequacy of the model, if a 'good' model, then its expected the residuals to be 'random' and 'close to zero' and mode validation is carried out by ACF plots. Clearly shows that there is no significant correlation between the residuals as all the points lie within the boundary lines.

The Ljung Box test examines the autocorrelation of residuals and if the autocorrelation is very small it is clear that the model adequately fit the data

H_0 : The data does not show lack of fit.

H_1 : The data show lack of fit

Box- Ljung Test

X-squared = 0.00059859, df = 1, p-value = 0.9805

Here p-value>0.05. So, we accept H_0 . i.e., residuals are uncorrelated. Plot ACF and PACF to find the correlation is presented in Figure 62.

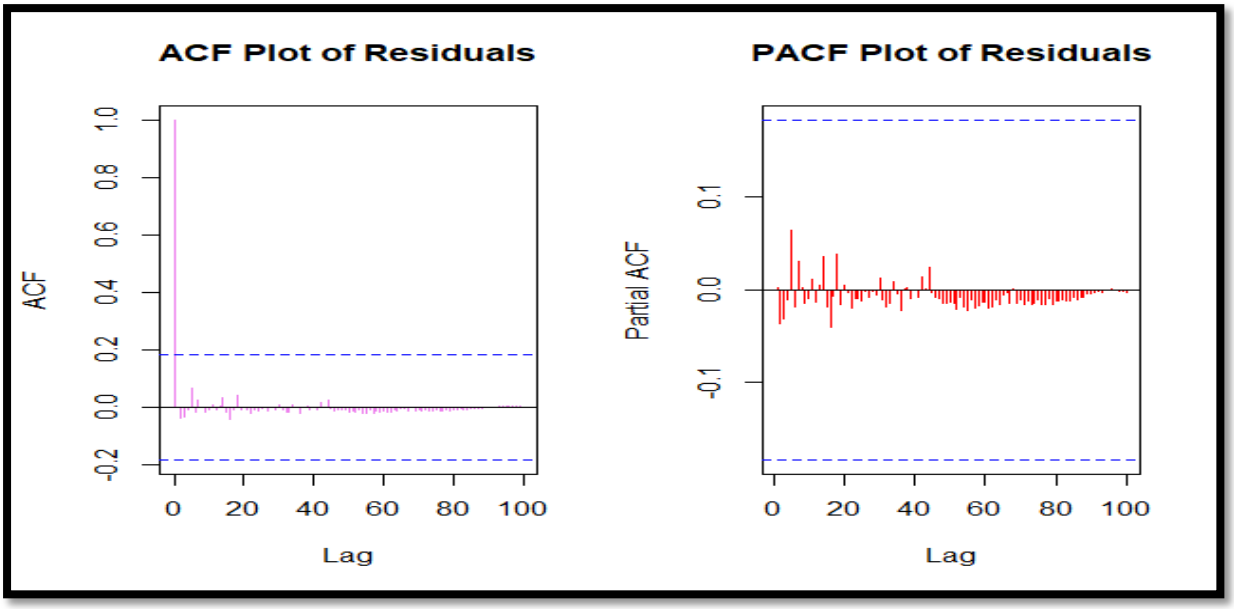


Figure 62: Plot for ACF and PACF of Residuals

To forecast the future mortality of COVID-19 use of ARIMA (0,1,1) model and values are interpreted in Table 29 and the forecasted plot is given as in Figure 63 indicating an increasing pattern of future demise.

Table 29: Forecasted values of COVID-19 deaths for consecutive weekends of three months (May 22-Aug 27)

Date	Forecast	80% Confidence Interval		95% Confidence Interval	
		Lower Limit	Upper Limit	Lower Limit	Upper Limit
July 4	463.3284	251.6410	675.0158	139.5804	787.0763
July 11	493.9833	275.4929	712.4737	159.8310	828.1355
July 18	524.6382	299.5503	749.7261	180.3959	868.8804
July 25	555.2931	323.7956	786.7905	201.2482	909.3379
Aug 1	585.9479	348.2137	823.6822	222.3647	949.5312
Aug 8	616.6028	372.7912	860.4144	243.7252	989.4805
Aug 15	647.2577	397.5167	896.9988	265.3117	1029.2037
Aug 22	677.9126	422.3796	933.4456	287.1086	1068.7166
Aug 29	708.5675	447.3710	969.7640	309.1019	1108.0331
Sept 5	739.2224	472.4826	1005.9622	331.2791	1147.1657
Sept 12	769.8773	497.7071	1042.0475	353.6288	1186.1257
Sept 19	800.5322	523.0378	1078.0266	376.1411	1224.9232
Sept 26	831.1871	548.4687	1113.9054	398.8067	1263.5674
Oct 3	861.8420	573.9945	1149.6894	421.6173	1302.0667
Oct 10	892.4968	599.6101	1185.3836	444.5652	1340.4285

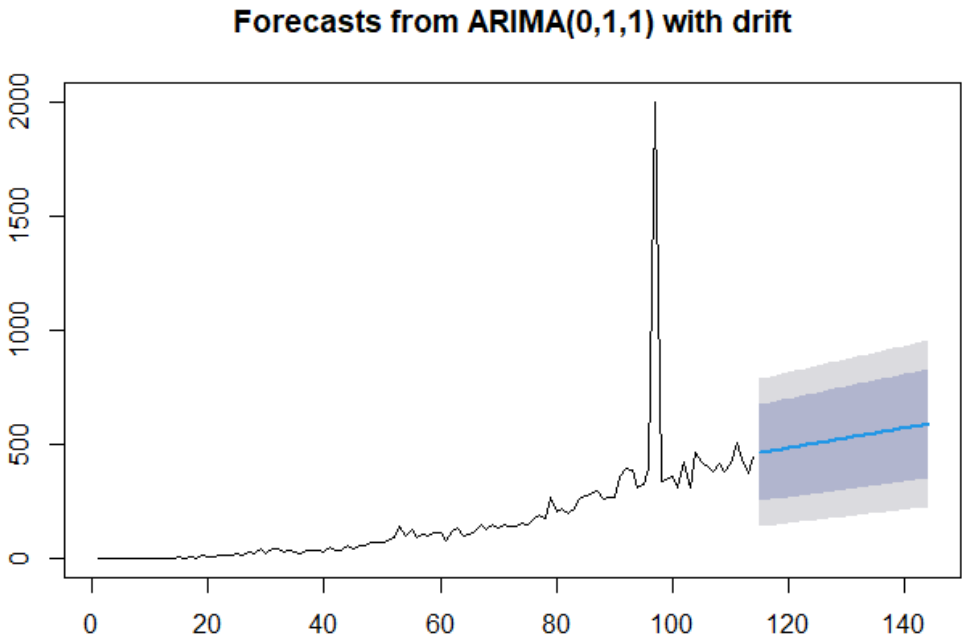


Figure 63: Forecast plot of COVID-19 mortality of India

3.2.3. Compartmental Models

Table 30: COVID-19 Cured, Deaths and Confirmed cases and active cases in south Indian states of India, as of May 19, 2021

States	Cured	Deaths	Confirmed	Active Cases (%)
Andaman and Nicobar Islands	6359	92	6674	223 (3.34)
Andhra Pradesh	1254291	9580	1475372	211501 (14.34)
Karnataka	1674487	22838	2272374	575049 (25.31)
Kerala	1846105	6612	2200706	347989 (15.81)
Lakshadweep	3915	15	5212	1282 (24.6)
Puducherry	69060	1212	87749	17477 (19.92)
Tamil Nadu	1403052	18369	1664350	242929 (14.6)
Telangana	485644	3012	536766	48110 (8.96)

Table 30 reveals that Karnataka and Lakshadweep exhibit a relatively higher percentage of active COVID-19 cases, both accounting for 25% of the total reported cases in the respective regions. Following closely is Puducherry, where active cases constitute 19.92% of the total, and Kerala, Andhra Pradesh, and Tamil Nadu with 15% each.

Table 31: Case Fatality Rate in south Indian states of India, as of May 19, 2021

State	Phase	Days	Minimum	Maximum	Mean	Std. Deviation
Andaman and Nicobar Islands	Phase1	183	0.508	4.313	1.568	0.578
	Resting phase	57	1.231	1.244	1.237	0.004
	Phase2	51	1.162	1.426	1.229	0.058
Andhra Pradesh	Phase1	178	0.813	2.879	1.493	0.721
	Resting phase	112	0.806	0.813	0.808	0.002
	Phase2	58	0.758	0.818	0.803	0.014
Karnataka	Phase1	307	1.310	37.500	4.228	5.364
	Resting phase	33	1.304	1.310	1.306	0.002
	Phase2	76	1.292	1.394	1.320	0.031
Kerala	Phase1	406	0.000	8.000	0.715	0.897
	Resting phase	13	0.414	0.419	0.416	0.002
	Phase2	52	0.357	0.432	0.414	0.020
Lakshadweep	Phase1	30	0.163	0.408	0.272	0.078

	Resting phase	18	0.134	0.160	0.143	0.007
	Phase2	35	0.075	0.394	0.238	0.114
Puducherry	Phase1	165	1.685	6.034	2.596	0.944
	Resting phase	112	1.664	1.691	1.676	0.007
	Phase2	65	1.616	1.730	1.668	0.028
Tamil Nadu	Phase1	265	0.000	50.000	3.713	7.174
	Resting phase	111	1.467	1.525	1.490	0.015
	Phase2	63	1.263	1.467	1.393	0.076
Telangana	Phase1	259	0.000	87.500	4.779	11.515
	Resting phase	97	0.547	0.578	0.553	0.007
	Phase2	68	0.549	0.634	0.568	0.021

Table 31 categorizes the duration of COVID-19 disease in various states, delineating the information into three distinct phases. Notably, each phase reflects a discernible variance in the average Case Fatality Rate (CFR). Across the majority of states, the average CFR remains below two. Nevertheless, a noteworthy observation is the alteration in the average CFR during phase one across almost all states. Subsequently, the resting phase and phase two exhibit a CFR consistently below 2, indicating a distinct pattern in the progression of the disease across different phases in the respective states.

Table 32: Parameter values of Regression Models for COVID-19 in states of South India, as of May 19, 2021

State	Phase	Model Summary					Parameter Estimates		
		R^2	F	df1	df2	Sig.	Constant	b1	b2
Andaman and Nicobar Islands	Phase 1	0.787	345.6	2	187	<0.001	0.1	-0.001	2.688E-06
	Phase 2	0.939	370.8	2	48	<0.001	-7746.9	35.594	-0.040
Andhra Pradesh	Phase 1	0.777	304.4	2	175	<0.001	-321699.4	4476.447	-12.535
	Phase 2	0.976	1121.9	2	55	<0.001	10397781.1	-55346.070	73.627
Karnataka	Phase 1	0.968	4906.6	2	324	<0.001	1.5	0.095	0.000
	Phase 2	0.975	1404.1	2	73	<0.001	27591727.5	-146626.706	194.737
Kerala	Phase 1	0.670	409.9	2	403	<0.001	-19787.8	293.836	-0.156
	Phase 2	0.927	311.6	2	49	<0.001	12011107.7	-62708.623	80.921
Lakshadweep	Phase 1	0.677	68.0	2	65	<0.001	-124.0	4.364	-0.016
	Phase 2	0.837	81.9	2	32	<0.001	-49073.5	677.552	-2.280
Puducherry	Phase 1	0.553	153.6	2	248	<0.001	-1320.1	33.095	-0.053
	Phase 2	0.996	7233.4	2	62	<0.001	750201.6	-4059.235	5.493
Tamil Nadu	Phase 1	0.745	382.9	2	262	<0.001	-24086.2	804.052	-2.379
	Phase 2	0.984	1852.4	2	60	<0.001	7462955.9	-39728.715	52.911
Telangana	Phase 1	0.747	378.8	2	256	<0.001	-6647.8	165.609	-0.124
	Phase 2	0.969	1010.9	2	65	<0.001	4814860.7	-25517.422	33.789

The Quadratic Regression Models applied to analyze the impact of COVID-19 across various states in India demonstrate the explanatory efficacy of these models, ranging from 55% during phase 1 to an impressive 99% during phase two in Puducherry. The comprehensive timeline of the COVID-19 pandemic has been systematically divided into three distinct phases, each characterized by fitting quadratic regression models. During the initial phase, Karnataka emerges as the state with the most optimal fit, closely followed by Andaman and Nicobar Islands.

In the second phase, Puducherry exhibits the best-fitted regression model, with Tamil Nadu and Andaman and Nicobar Islands trailing behind. Notably, all states exhibit a linear trend with a subtle slope toward the x-axis during the resting phase, emphasizing the consistency of this pattern across the regions. Additionally, the R-square values exceeding 90% in all states underscore the robustness of the quadratic regression models in capturing the underlying trends during the resting phase (Table 32).

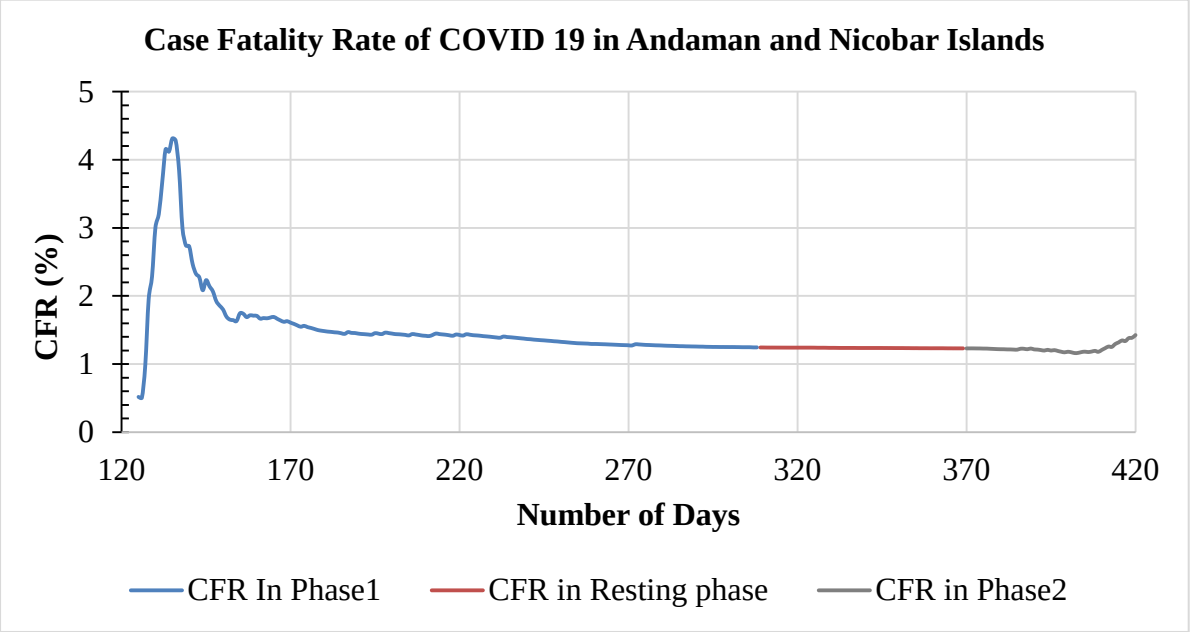


Figure 64: CFR of COVID-19 in Andaman and Nicobar Islands

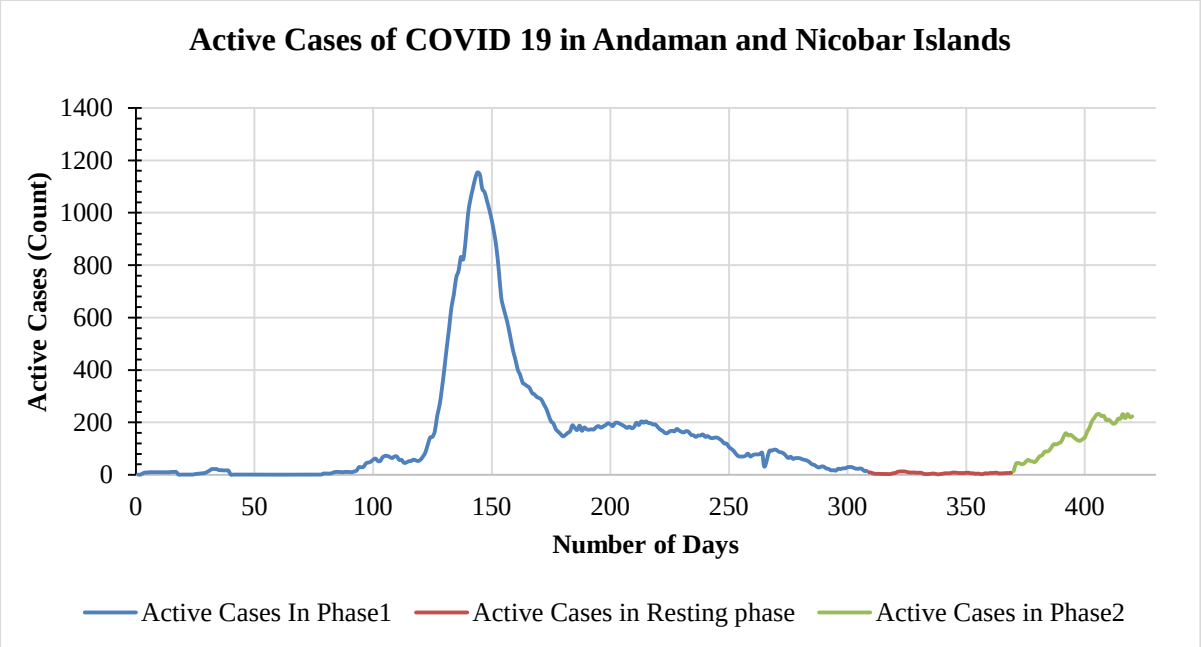


Figure 65: Active cases of COVID-19 in Andaman and Nicobar Islands

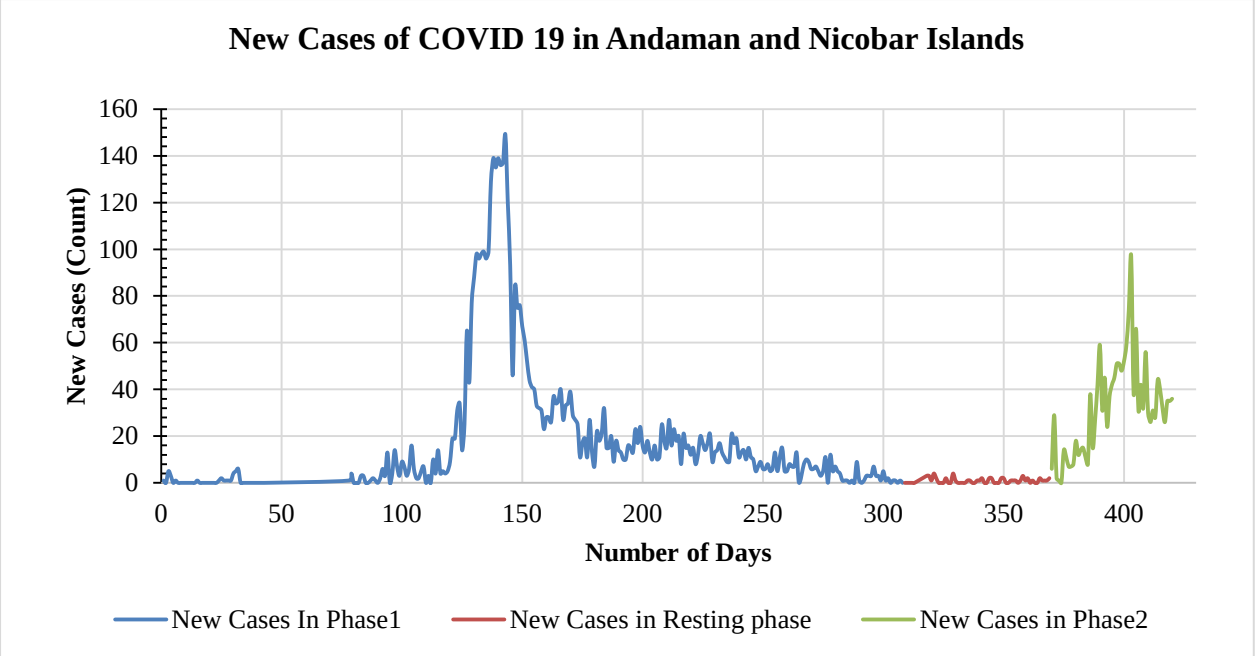


Figure 66: New cases of COVID-19 in Andaman and Nicobar Islands

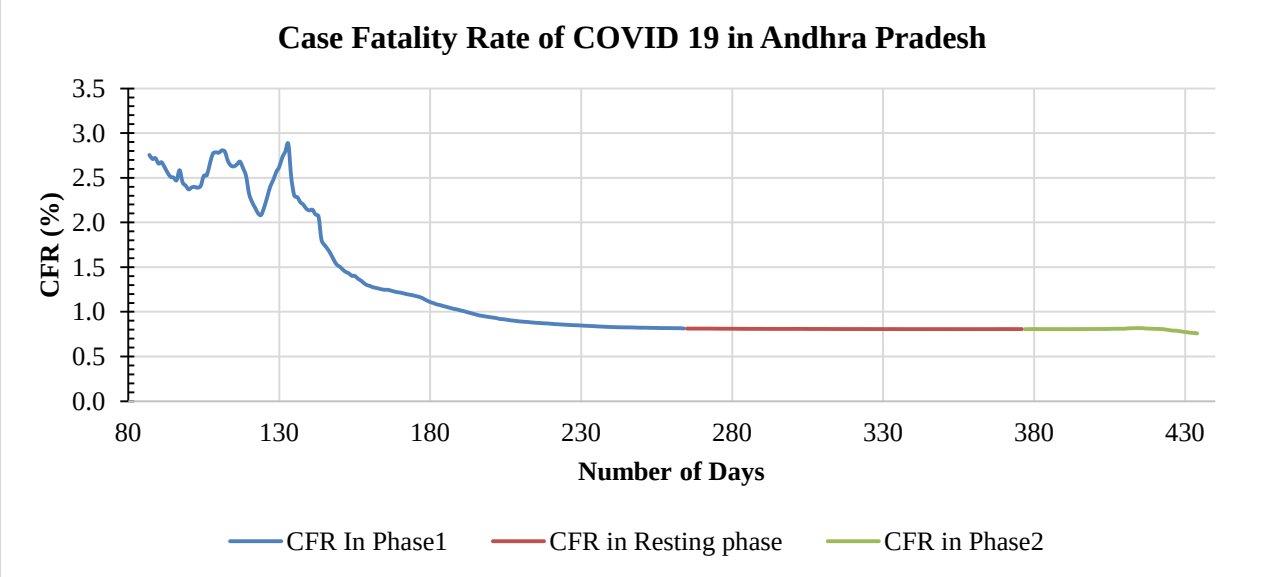


Figure 67: CFR of COVID-19 in Andhra Pradesh

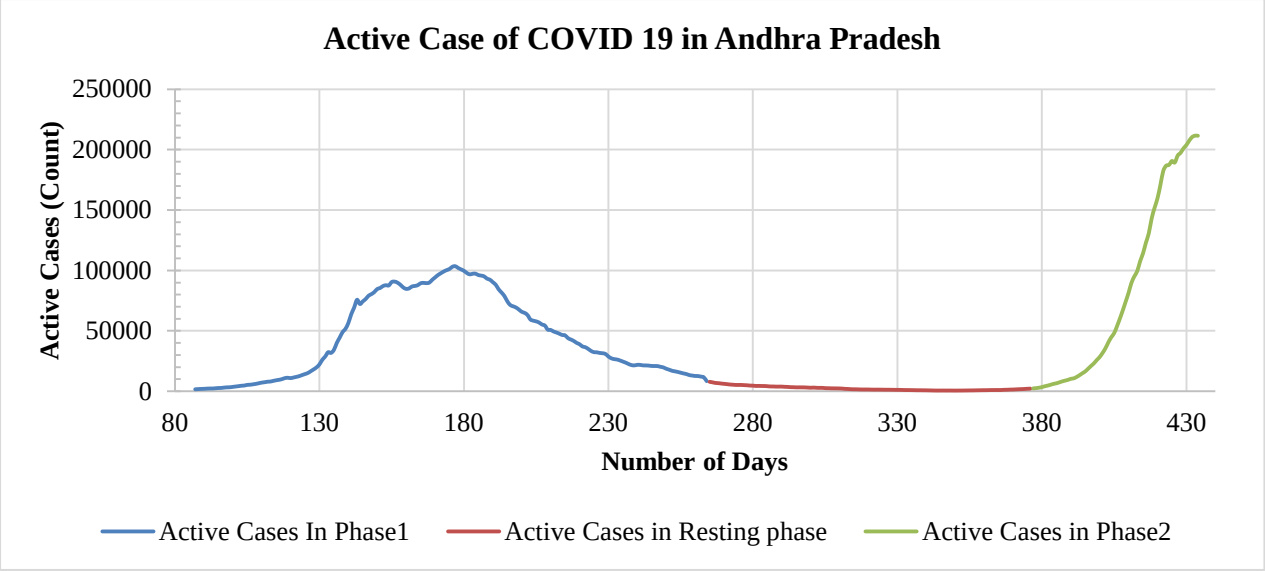


Figure 68: Active cases of COVID-19 in Andhra Pradesh

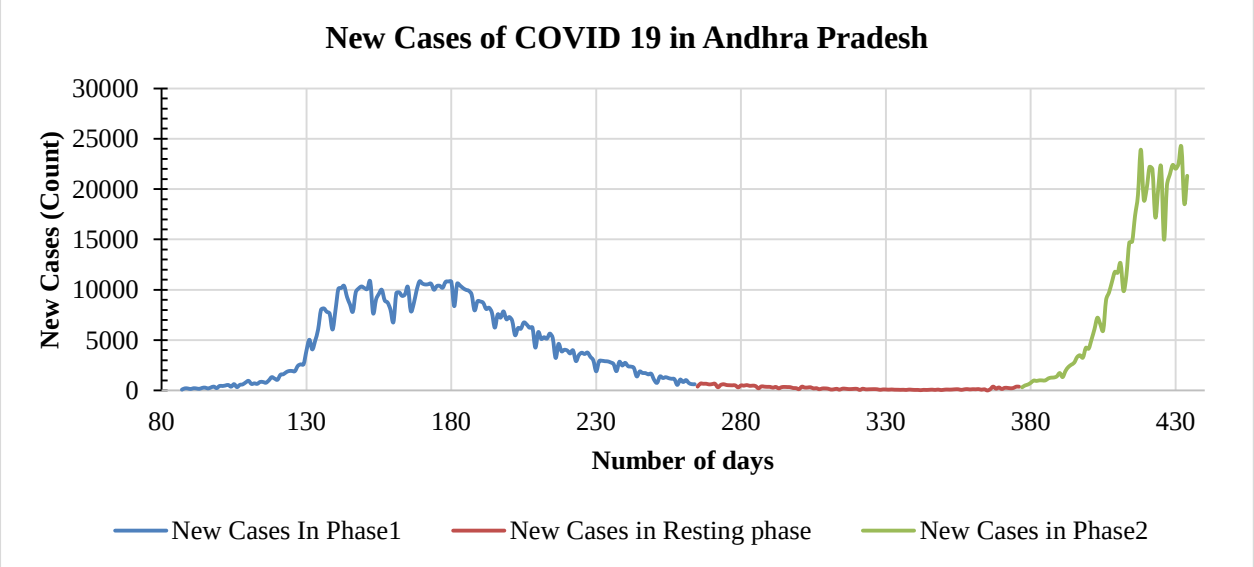


Figure 69: New cases of COVID-19 in Andhra Pradesh

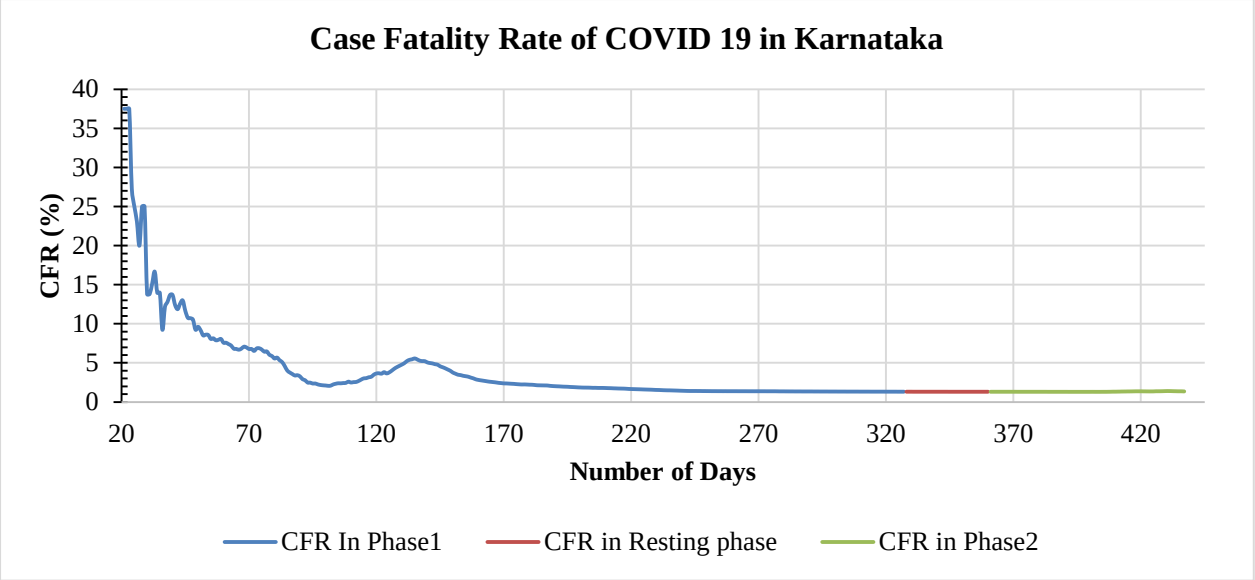


Figure 70: CFR of COVID-19 in Karnataka

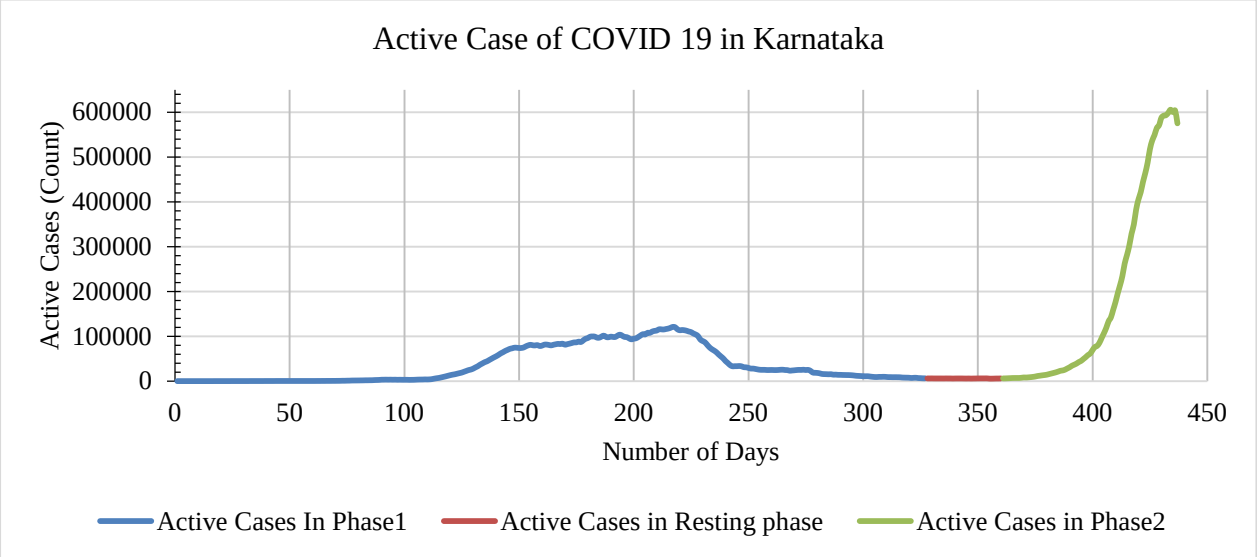


Figure 71: Active cases of COVID-19 in Karnataka

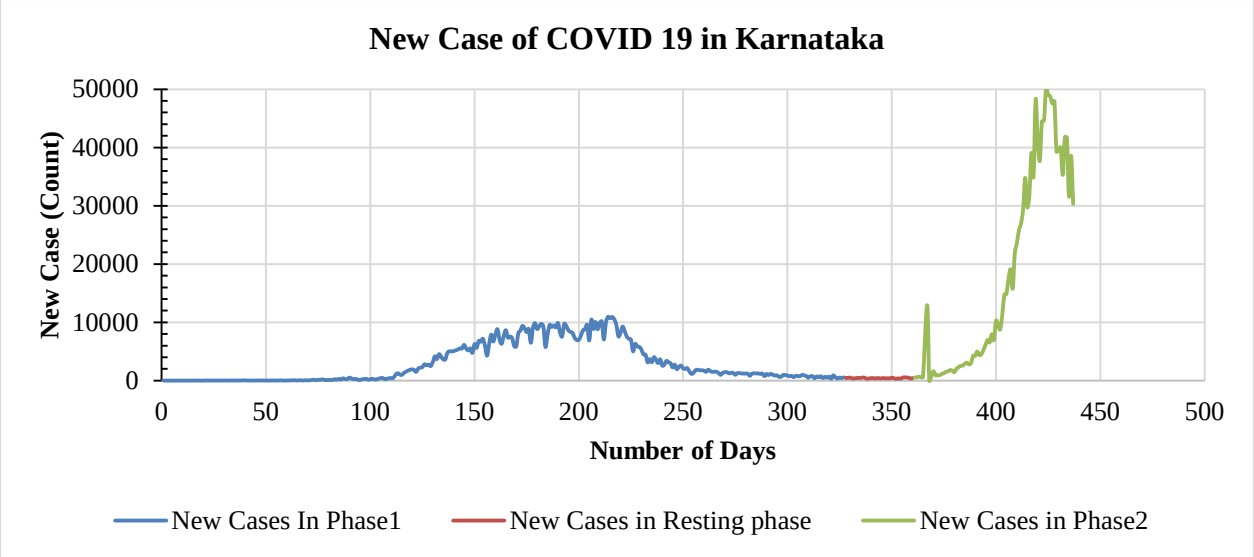


Figure 72: New cases of COVID-19 in Karnataka

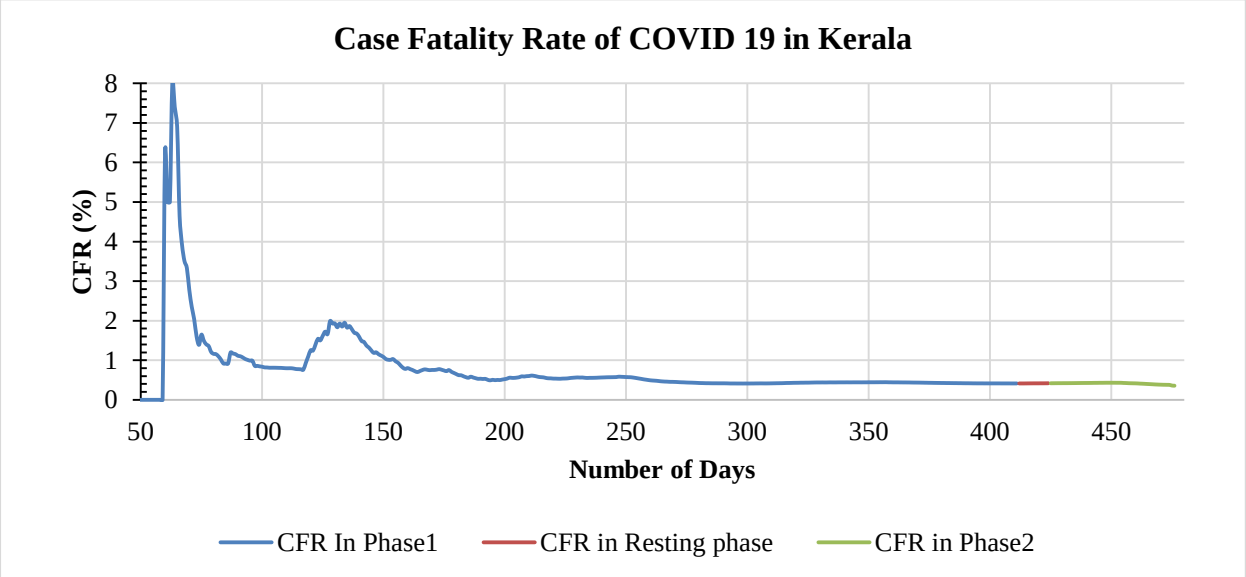


Figure 73: CFR of COVID-19 in Kerala

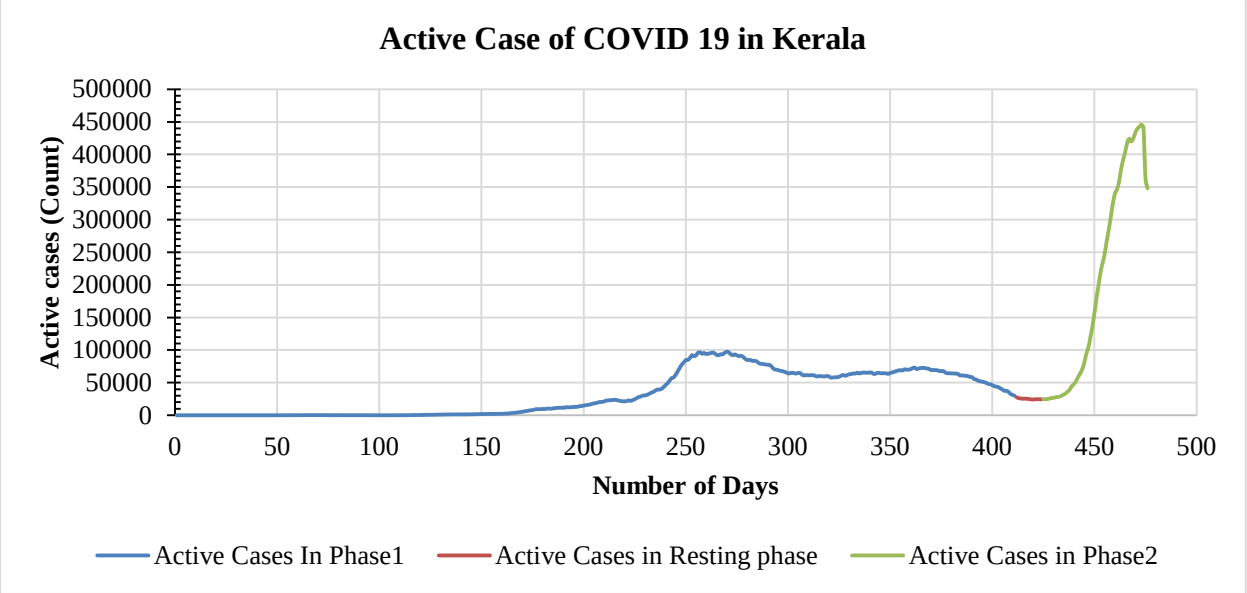


Figure 74: Active cases of COVID-19 in Kerala

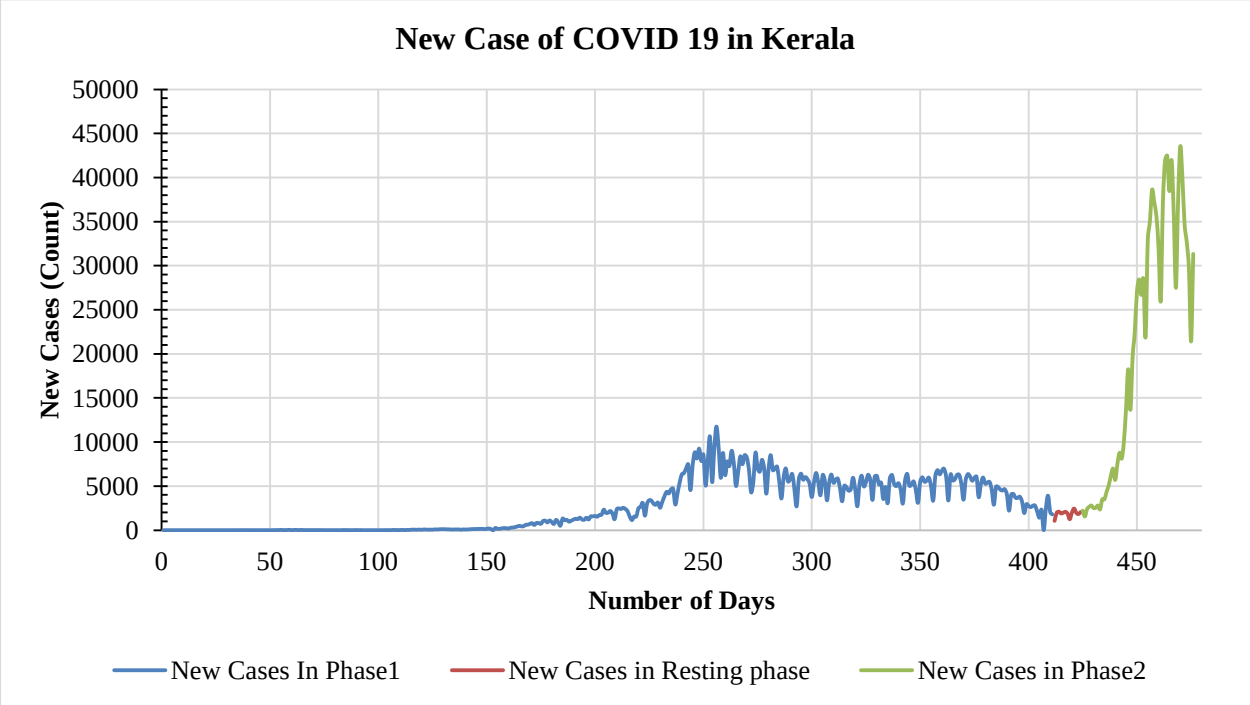


Figure 75: New cases of COVID-19 in Kerala

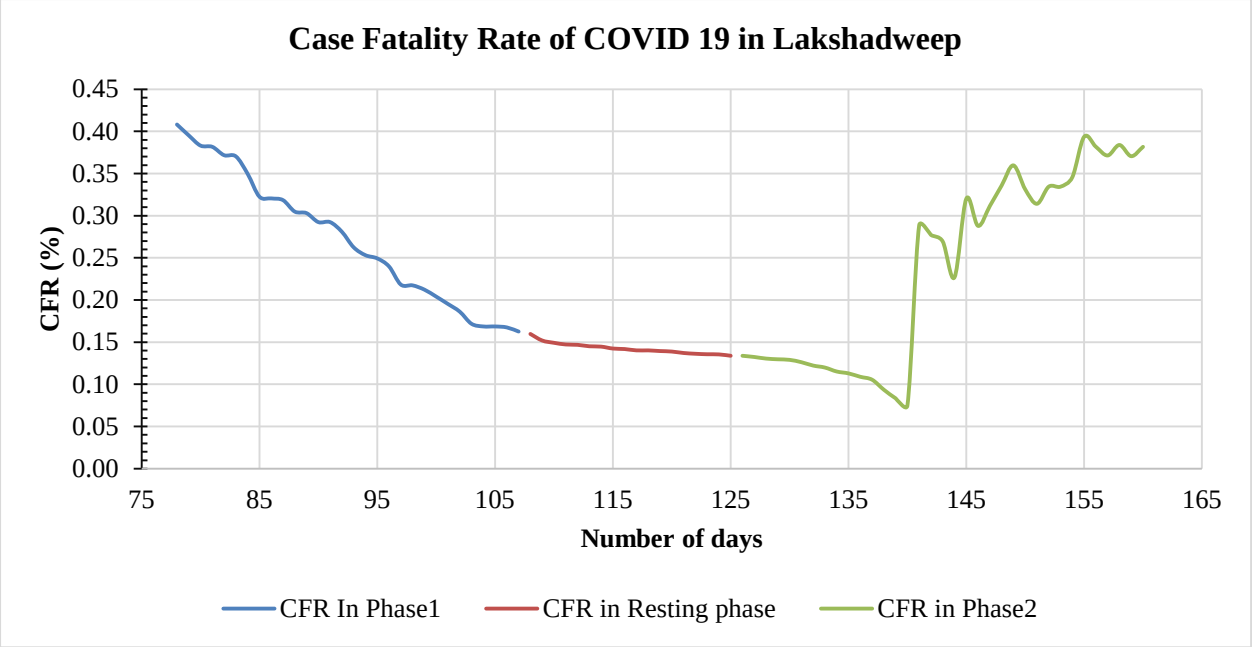


Figure 76: CFR of COVID-19 in Lakshadweep

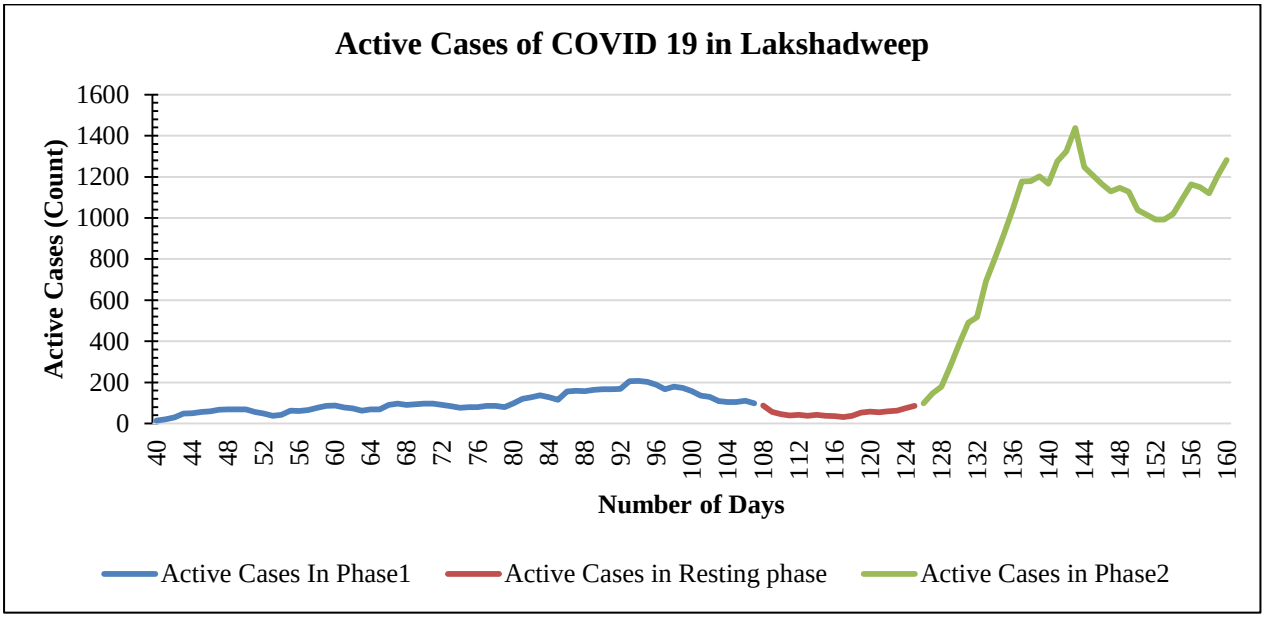


Figure 77: Active cases of COVID-19 in Lakshadweep

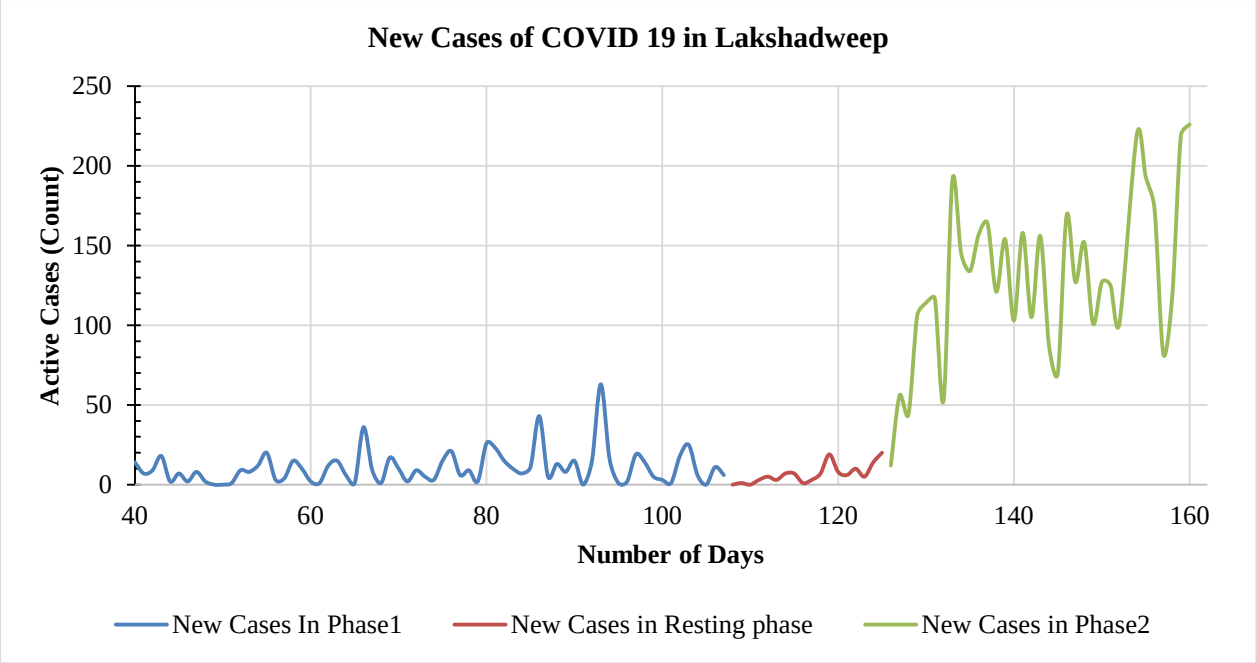


Figure 78: New cases of COVID-19 in Lakshadweep

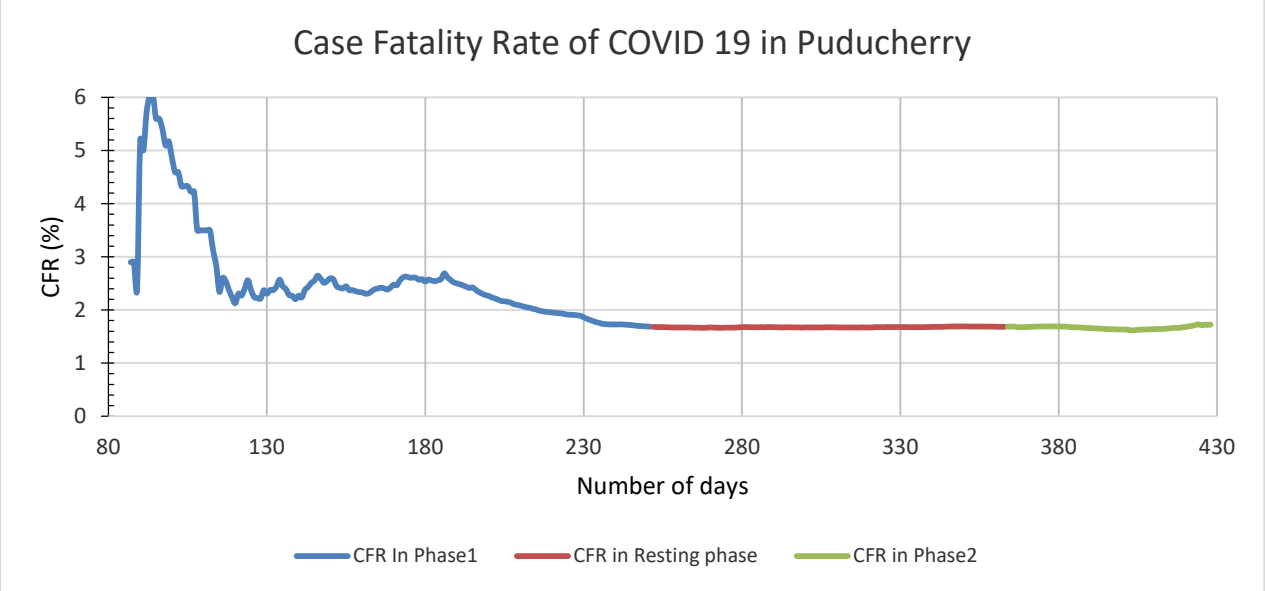


Figure 79: CFR of COVID-19 in Puducherry

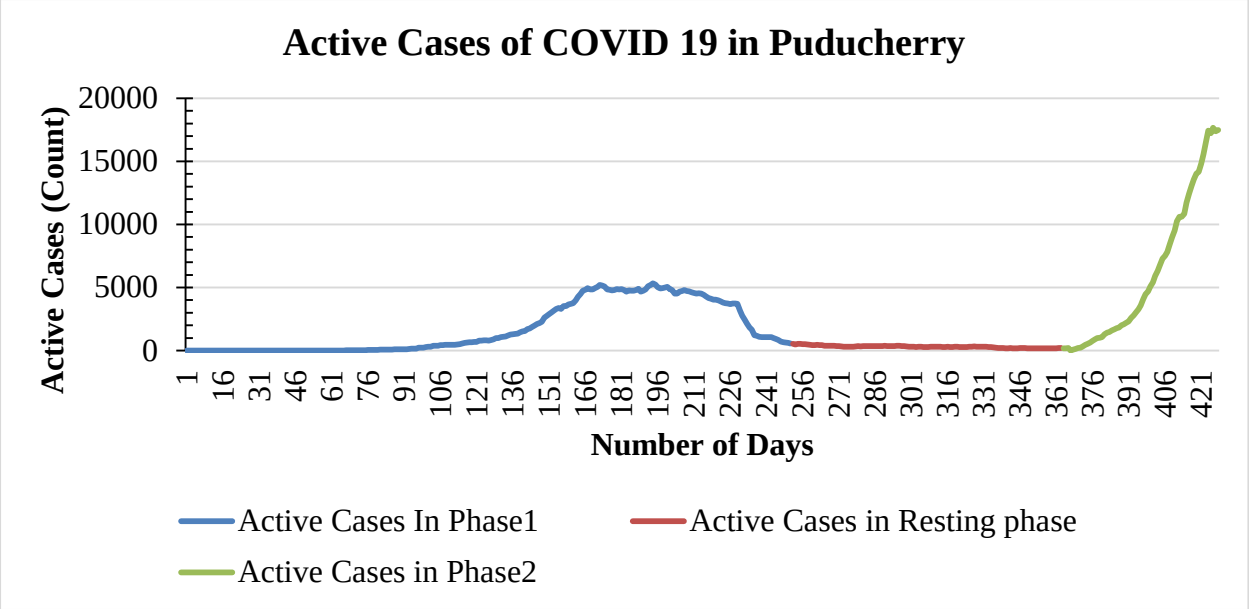


Figure 80: Active cases of COVID-19 in Puducherry

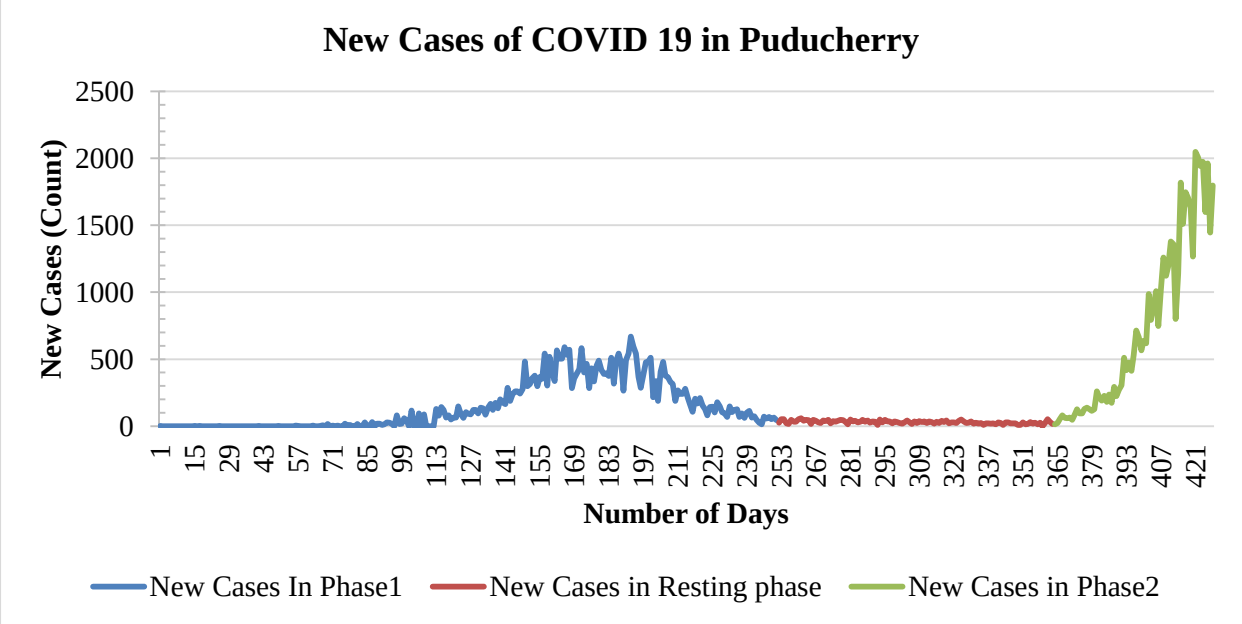


Figure 81: New cases of COVID-19 in Puducherry

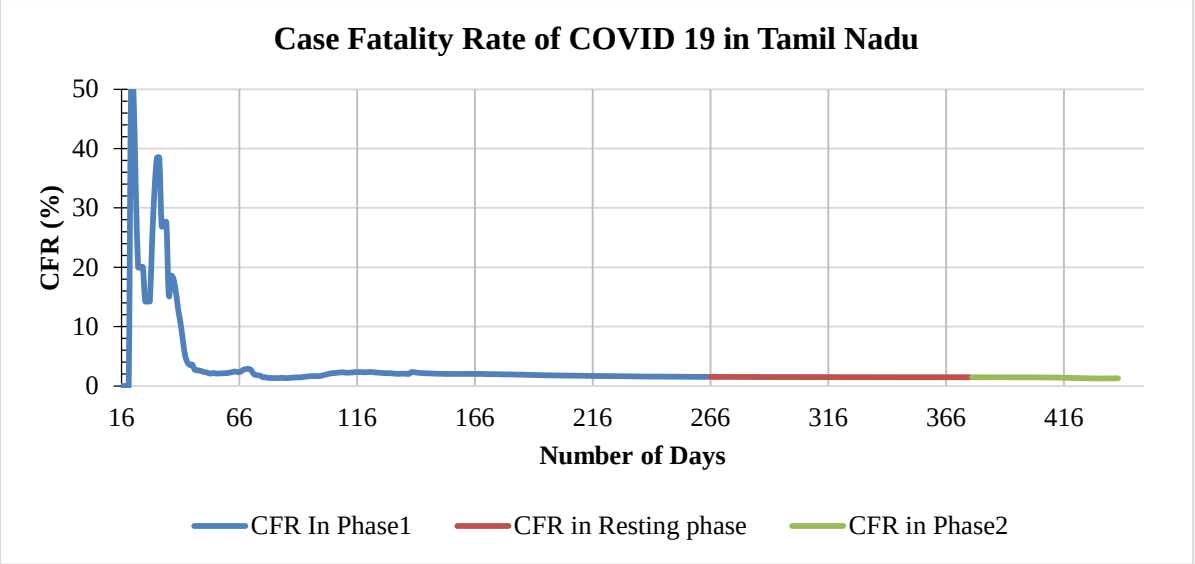


Figure 82: CFR of COVID-19 in Tamil Nadu

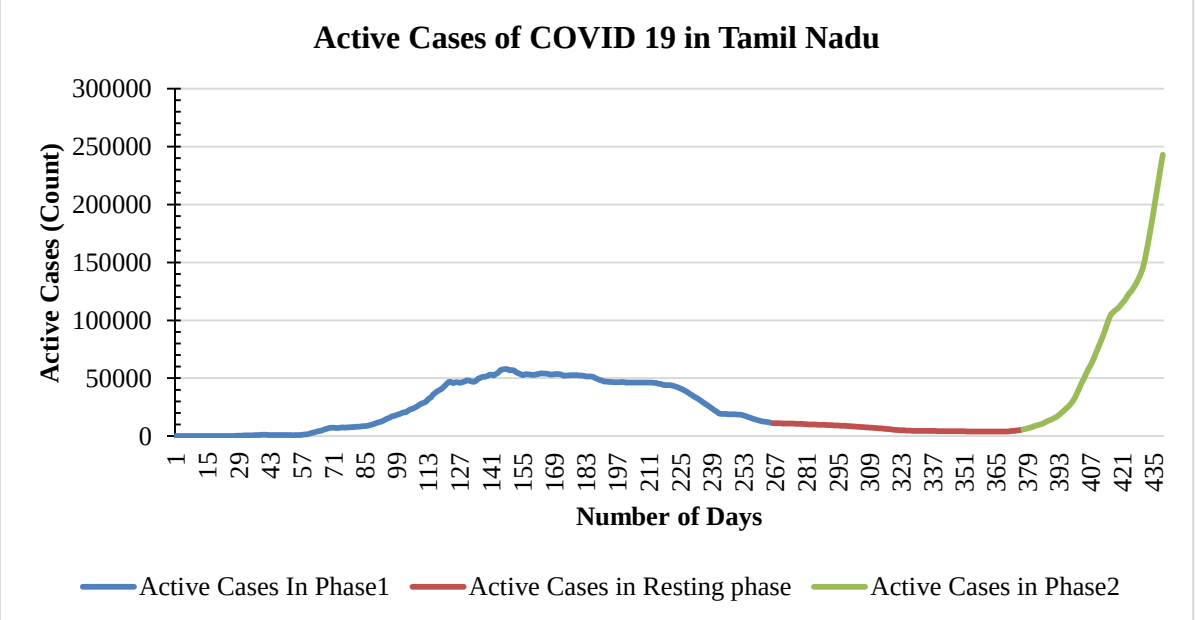


Figure 83: Active cases of COVID-19 in Tamil Nadu

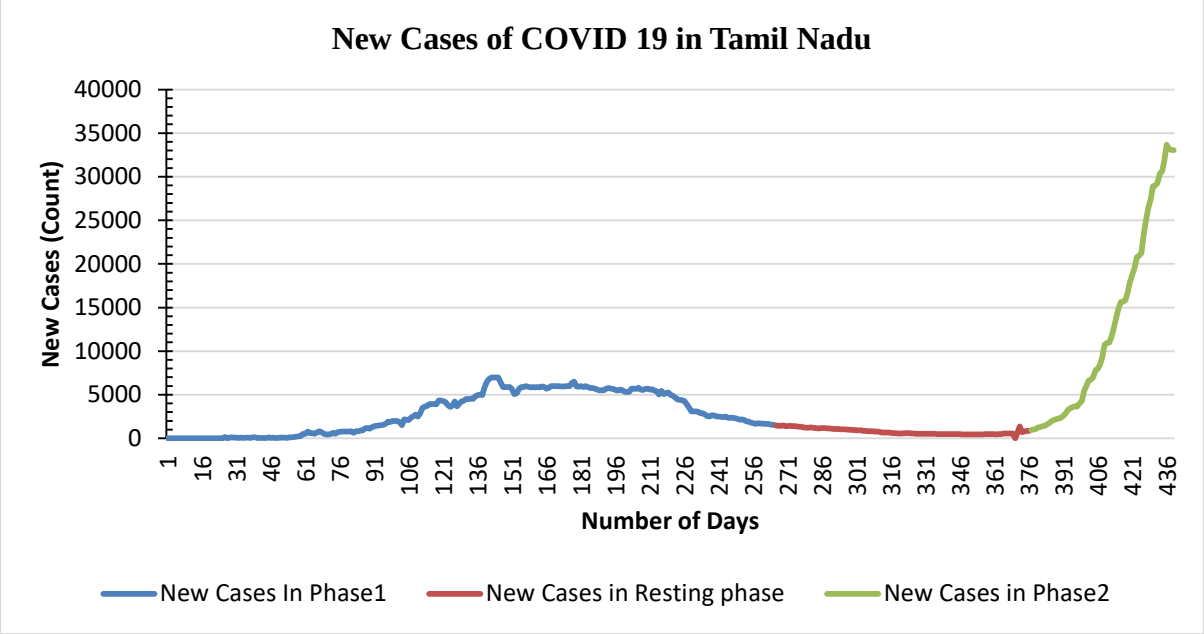


Figure 84: New cases of COVID-19 in Tamil Nadu

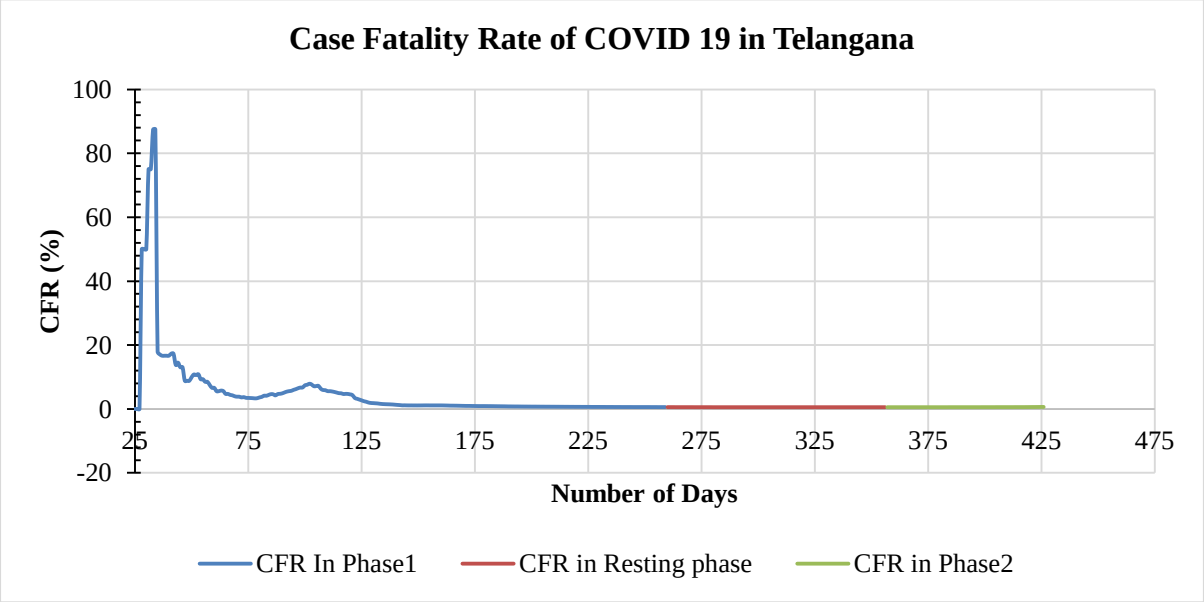


Figure 85: CFR of COVID-19 in Telangana

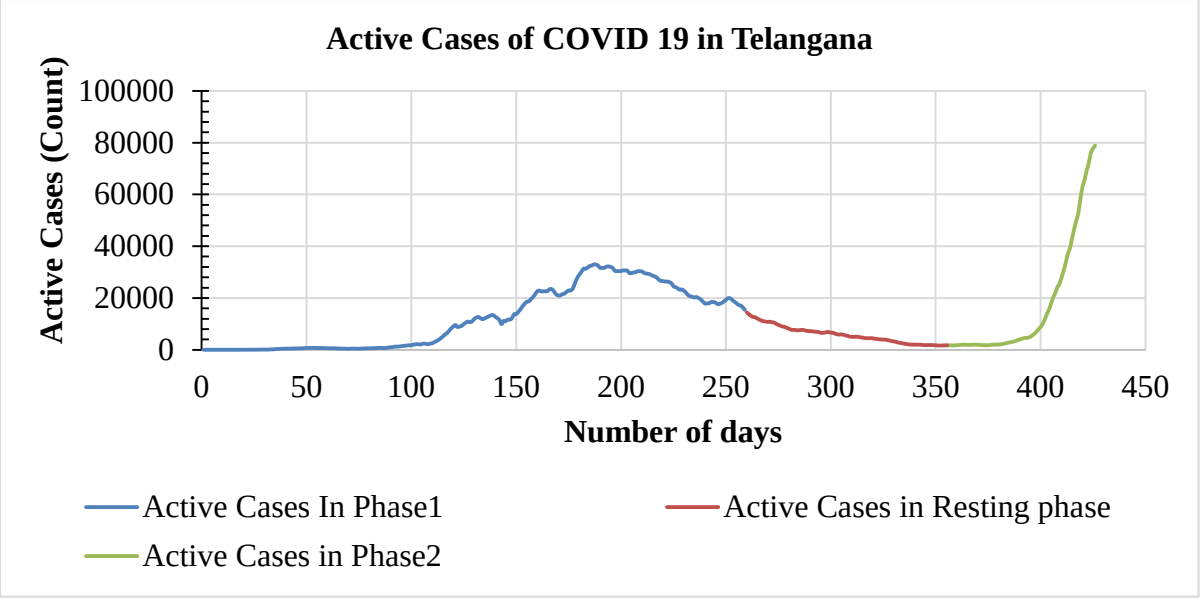


Figure 86: Active Cases of COVID-19 in Telangana

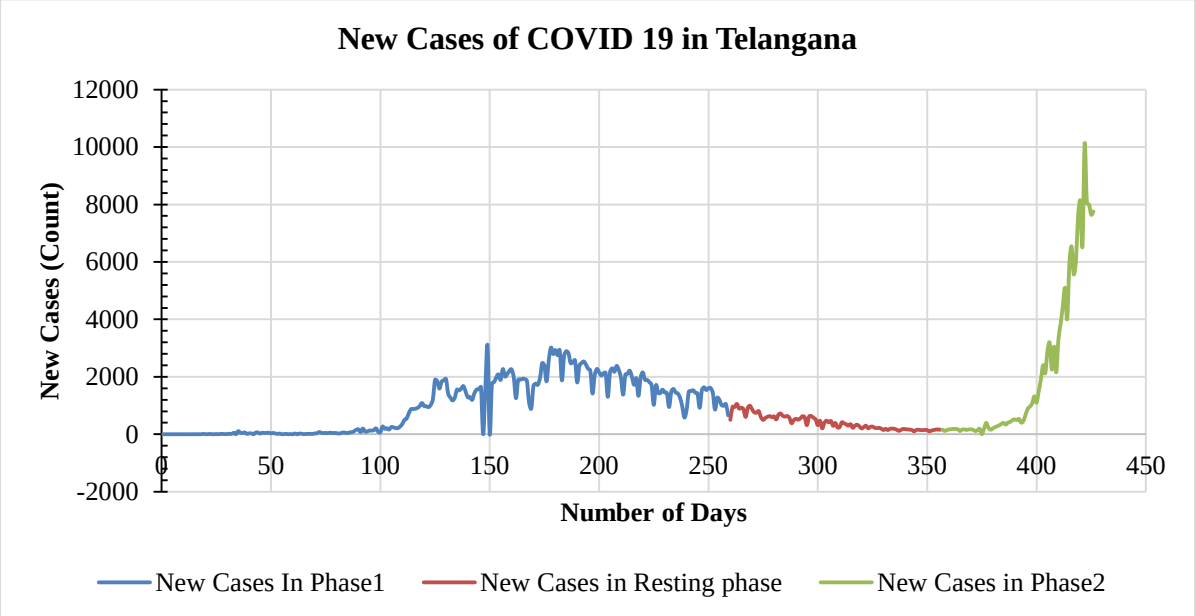


Figure 87: New Cases of COVID-19 in Telangana

In the early phase, the case fatality rate (CFR) in South Indian states and union territories showed fluctuations before stabilizing between 0.5 to 2. Notable stabilization points included the 160th and 180th days in Andaman and Nicobar Islands and Andhra Pradesh (Figure 64, 67), respectively. Karnataka and Kerala stabilized within 70 to 100 days (Figure 70, 73), while Lakshadweep had a less pronounced shift toward a linear pattern (Figure 76). During the first and resting phases, Lakshadweep exhibited a decreasing trend below 0.1, followed by an increase up to 0.4. Puducherry's CFR stabilized around 1.8 by the end of the first phase, persisting through subsequent phases (Figure 79). Tamil Nadu saw a drastic CFR increase in Phase 1, stabilizing by the 66th day (Figure 82). Telangana showed a trend similar to Karnataka and Kerala, stabilizing by the 75th day with minor fluctuations (Figure 85).

New and active COVID-19 cases showed similar trends across South Indian states. In Andaman and Nicobar Islands, a notable surge around the 110th day peaked at approximately 1180 active cases and 150 new cases by the 150th day, followed by a gradual decline (Figure 65, 66). Phase two, starting around the 309th day, didn't reach the previous peak.

In Andhra Pradesh, the second phase's peak (around 0.22 million active cases and 25 thousand new cases) doubled that of the first (around 0.1 million active cases and 12 thousand new cases). Phase one peaked by the 180th day, transitioning to a resting phase by the 265th day, with phase two beginning on the 377th day (Figure 68, 69).

Karnataka saw a fivefold increase in active and new cases during phase two compared to phase one. Phase one surged dramatically after the 100th day, reaching a peak by the 250th day, followed by a brief resting period by the 328th day. The surge resumed,

reaching its maximum by the 425th day (around 0.6 million active cases and 50 thousand new cases) (Figure 71, 72). Similarly, the Figure 77 and Figure 78 present the active and new cases during phase two compared to phase one for Lakshadweep.

States like Kerala (Figure 74, 75), Puducherry (Figure 80, 81), Tamil Nadu (Figure 83, 84), and Telangana (Figure 86, 87) mirrored this fivefold increase. Kerala, reporting the first cases in India, had a stable initial quarter, with a trigger around the 200th day. Puducherry, Tamil Nadu, and Telangana triggered around the 100th day, with shorter resting periods. The second phase witnessed a drastic increase, reaching 4.5 to 5 times the peak in both active and new cases.

3.2.4. Spatial Epidemiology and Forecasting

The section described the results of ARIMA modelling, correlation analysis, and exploratory analysis.

ARIMA

India

The time series plot displayed an increasing trend for the daily counts of cases in India. ACF and PACF were plotted to examine the stationarity of the data. High auto correlation was found till lag 10 (Figure 88). The lag 1 partial auto correlation was also positive. This indicated the need for a higher order non-seasonal differencing.

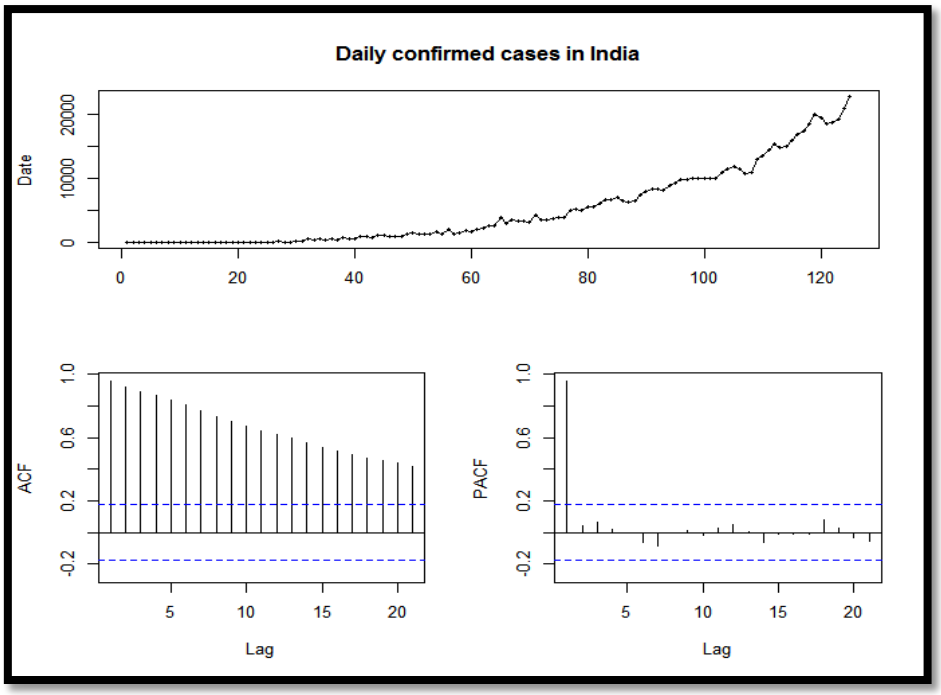


Figure 88: ACF and PACF plots of daily confirmed COVID-19 cases in India

ADF test was performed to confirm the non-stationarity of the data. Table 32 gives the results of ADF test. A non-significant *p* value = 0.99 at 95% CI was obtained.

Table 33: Augmented Dickey-Fuller Test Results for Daily Confirmed Cases in India

Augmented Dickey-Fuller test	
Data: Daily counts of Confirmed Cases in India	
Lag Order	4
p-value	0.99
Alternative Hypothesis: stationary	

The data was differenced for d=1.

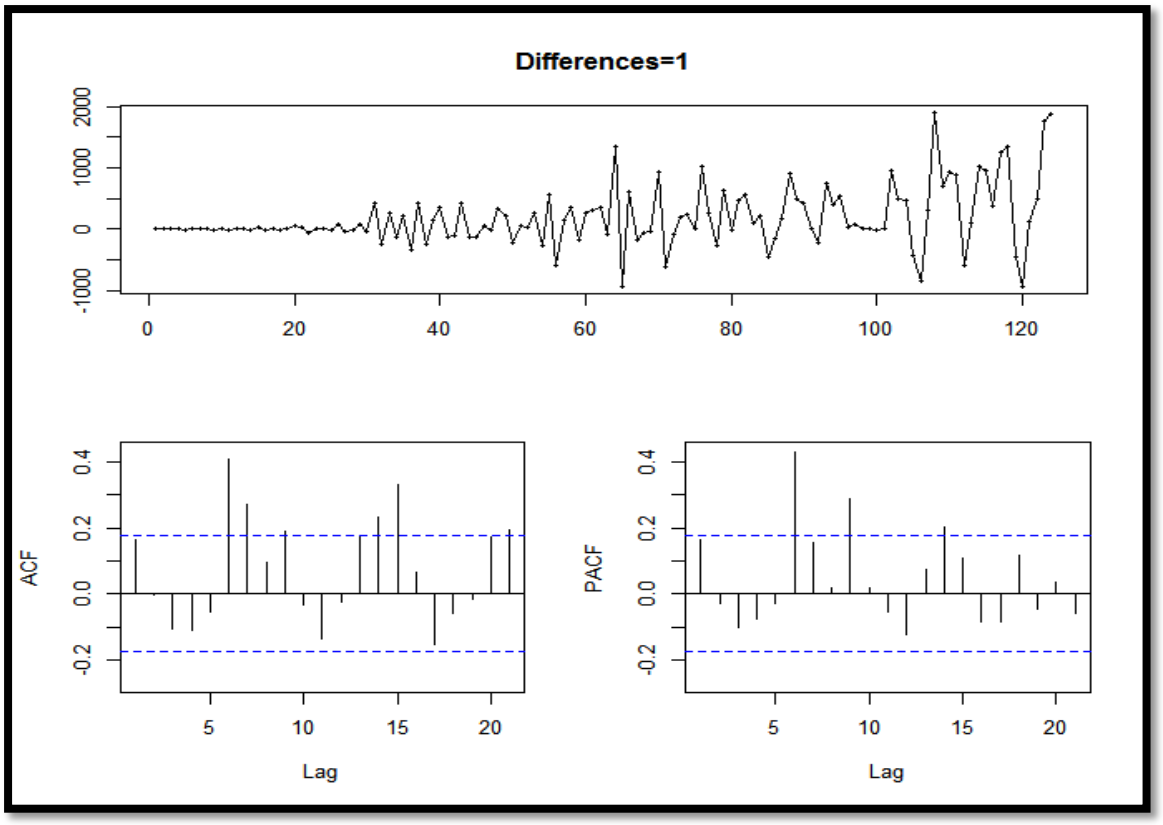


Figure 89: ACF and PACF plots of the differenced data

The time series plot showed no trend (Figure 89). The lag1 auto correlation was positive. The data seemed to be stationary. ADF test was again performed to conclude the result.

Table 34: Augmented Dickey-Fuller Test Results for Differenced Daily Confirmed Cases in India

Augmented Dickey-Fuller test	
Data: Daily counts of Confirmed Cases in India	
Lag Order	4
p-value	0.01
Alternative Hypothesis: stationary	

A significant *p* value at 95% CI was obtained. Data was confirmed to be stationary (Table 34). However, a pattern was found in the ACF of differenced data. Different number of AR and MA terms were added to remove the auto correlation pattern. The model ARIMA (9,1,9) with lowest AIC value of 1838.1 was selected to be the best model. The following table 35 gives the list of competed models and their AIC values.

Table 35: AIC values of competed models

Model	AIC
ARIMA (9,1,7)	1843.57
ARIMA (9,1,9)	**1838.1**
ARIMA (9,1,8)	1841.47

ARIMA (9,1,9) was fitted for the data of daily counts of confirmed COVID-19 cases in India. Residuals were plotted and their ACF and PACF drawn (Figure 90). Box-Ljung test for independence was performed to examine the independence of residuals (Table 36). The test resulted in a non-significant *p* value=0.6681 at 95% CI. The residuals were confirmed to be white noise. The following table and figure gives the test results and residual plots (Figure 90).

Table 36: Box-Ljung Test for Residuals from ARIMA (9,1,9) in India

Box-Ljung test	
Data: Residuals from ARIMA (9,1,9)	
X-squared	0.18382
Df	1
p value	0.6681

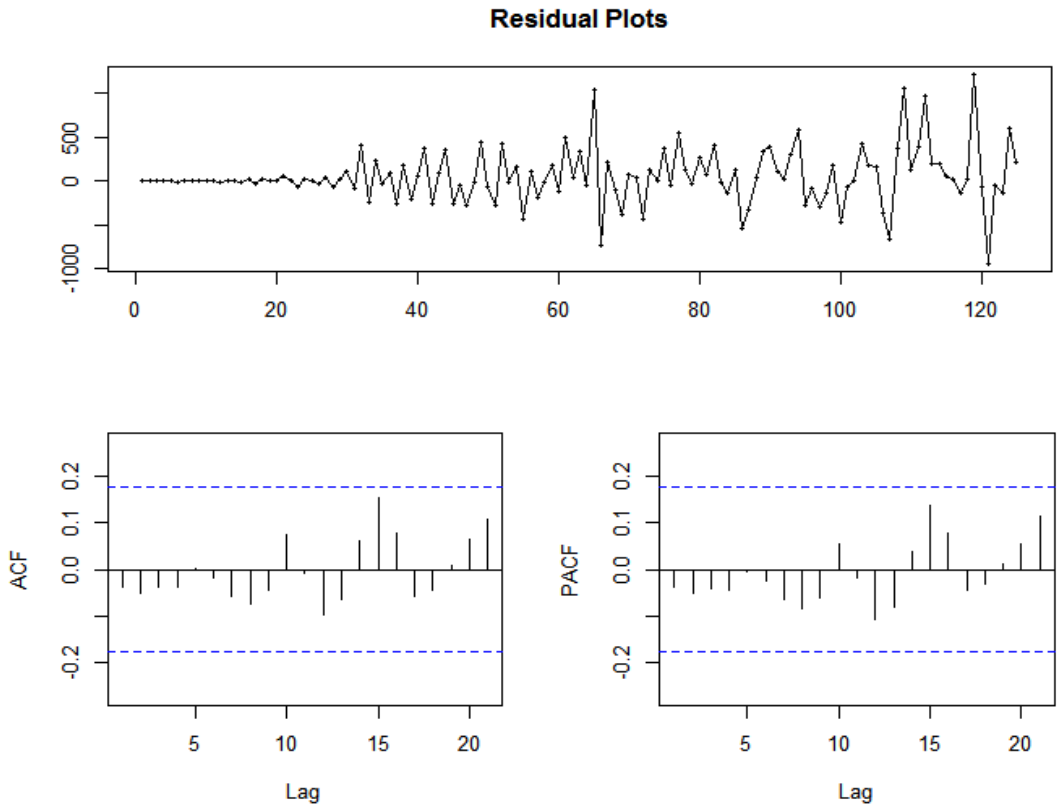


Figure 90: Residual plots of ARIMA (9,1,9) for the data of daily confirmed cases in India.

The fitted model equation is;

$$\Delta z_{t(India)} = C - 0.7030\ \Delta x_{t-1} - 0.1522\ \Delta x_{t-2} + 0.032\Delta x_{t-3} - 0.1477\ \Delta x_{t-4} - 0.0637\ \Delta x_{t-5} + 0.3183\ \Delta x_{t-6} + 0.6727\ \Delta x_{t-7} + 0.402\ \Delta x_{t-8} + 0.4946\ \Delta x_{t-9} + 0.7642\varepsilon_{t-1} + 0.2420\varepsilon_{t-2} - 0.1702\varepsilon_{t-3} + 0.0093\varepsilon_{t-4} - 0.0538\varepsilon_{t-5} + 0.3227\varepsilon_{t-6} + 0.169\varepsilon_{t-7} - 0.3191\varepsilon_{t-8} - 0.8093\varepsilon_{t-9} + \varepsilon_t$$

Interval (95% PI) as well as point forecasts were made for India. Figure 91 gives the plot of 15-day forecast in India. Figure 91,92, and 99 gives the point and interval forecast results in India, The USA and Italy respectively.

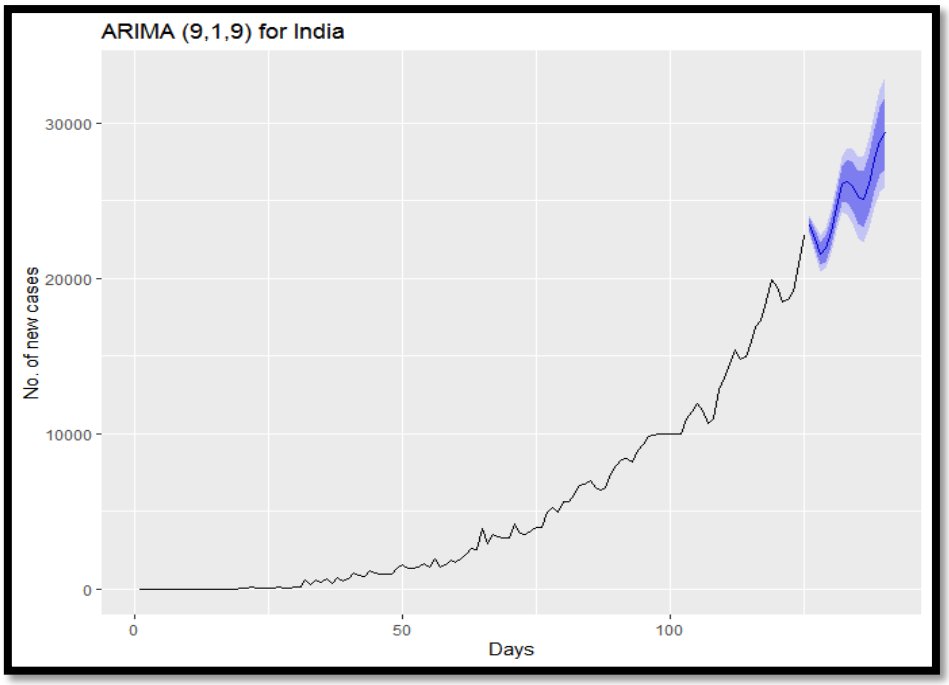


Figure 91: A 15-day forecast for daily confirmed cases in India

The USA

A linear trend was examined for the time series plot of the USA. ACF and PACF plots (Figure 92) showed significant auto correlation even beyond lag 16.

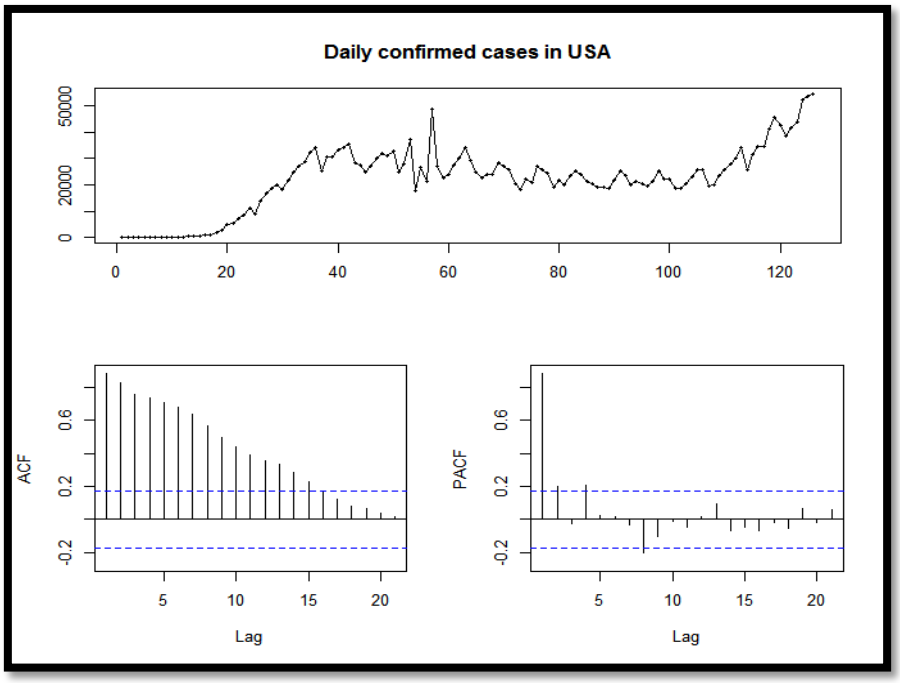


Figure 92: Time series, ACF and PACF plots of daily confirmed COVID-19 cases in The USA

To confirm the findings, ADF test was performed. The test showed non stationarity in the data with non-significant *p* value of 0.9588 with 95% confidence (Table 37). Thus, data proved to be non-stationary.

As a next step, the data was differenced for d=1 to stabilise the mean of daily confirmed COVID-19 cases in The USA. The trends in data were eliminated after first order differencing. The following Figure 93 gives the time series plot, ACF, PACF plots of the differenced data.

Table 37: ADF test results for the USA

Augmented Dickey-Fuller test	
Data: Daily counts of Confirmed Cases in The USA	
Dickey-Fuller	-0.80491
Lag order	4
p-value	0.9588
Alternative Hypothesis: stationary	

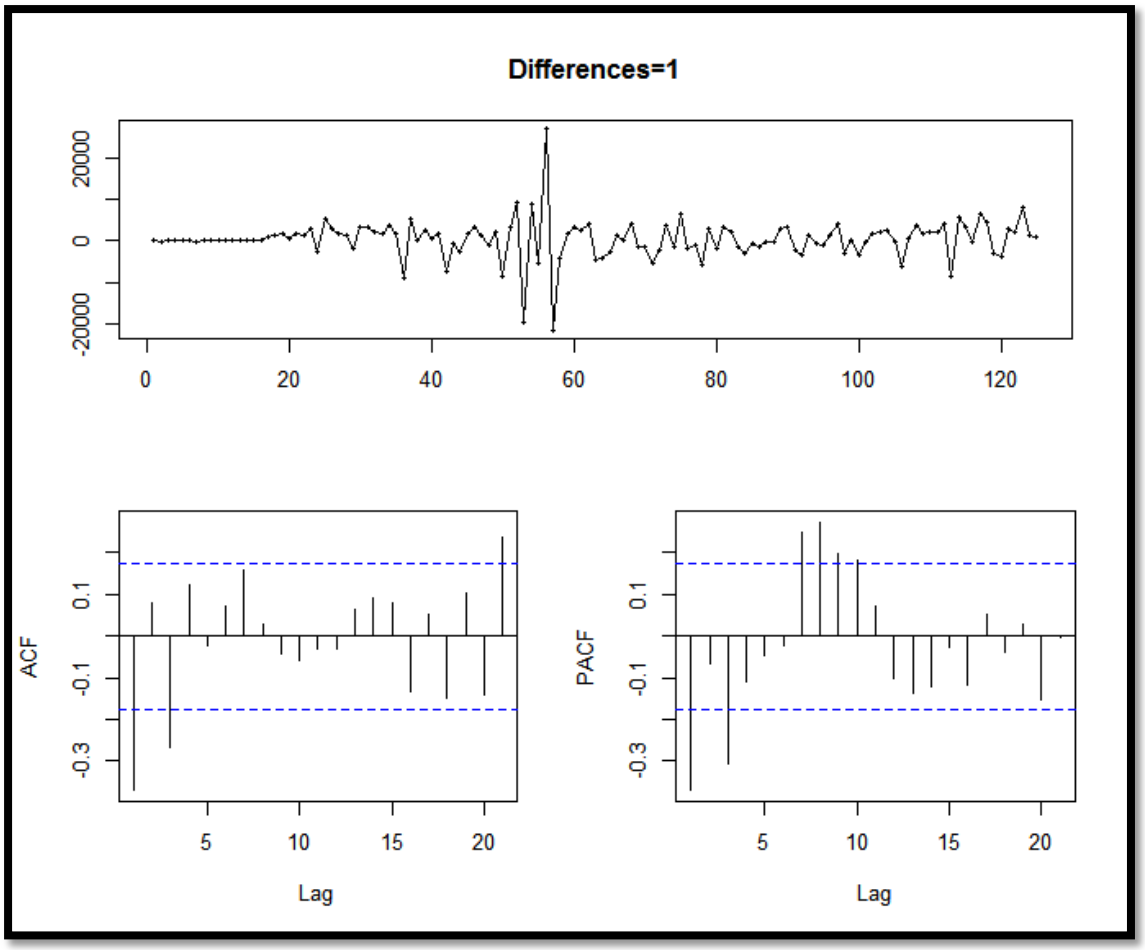


Figure 93: The Time Series plot, ACF and PACF plots of first order differenced data of daily number of confirmed cases in The USA.

The time series plot was observed to be stationary. The lag 1 auto correlation was low. However, there were significant spikes in the auto correlation and partial auto correlation plots. The stationarity of the data was confirmed with ACF test. Test gave a significant *p* value = 0.01 with 95% confidence thereby indicating stationarity (Table 38).

Table 38: ADF test for Differenced Daily Counts of Confirmed Cases in The USA

Augmented Dickey-Fuller test	
Data: Daily counts of Confirmed Cases in The USA	
Dickey-Fuller	-6.7108
Lag order	4
p-value	0.01
Alternative Hypothesis: stationary	

Several ARIMA models were developed and the model with the lowest AIC value was adopted as the best model. The following table gives the AIC values of the competed models. ARIMA (7,1,2) was chosen to be the best model with AIC value 2426.72 (Table 39).

Table 39: Comparison of ARIMA Models for USA Data Using AIC Values

Model (USA)	**AIC**
ARIMA (4,1,3)	2428.6
ARIMA (7,1,2)	**2426.72**
ARIMA (6,1,5)	2431.54
ARIMA (2,1,2)	1392.9
ARIMA (6,1,2)	1375.1
ARIMA (7,1,2)	1375.9

ARIMA (7,1,2) was fitted and residuals were diagnosed for white noise. The following figure gives residual plots of the fitted model (Figure 94).

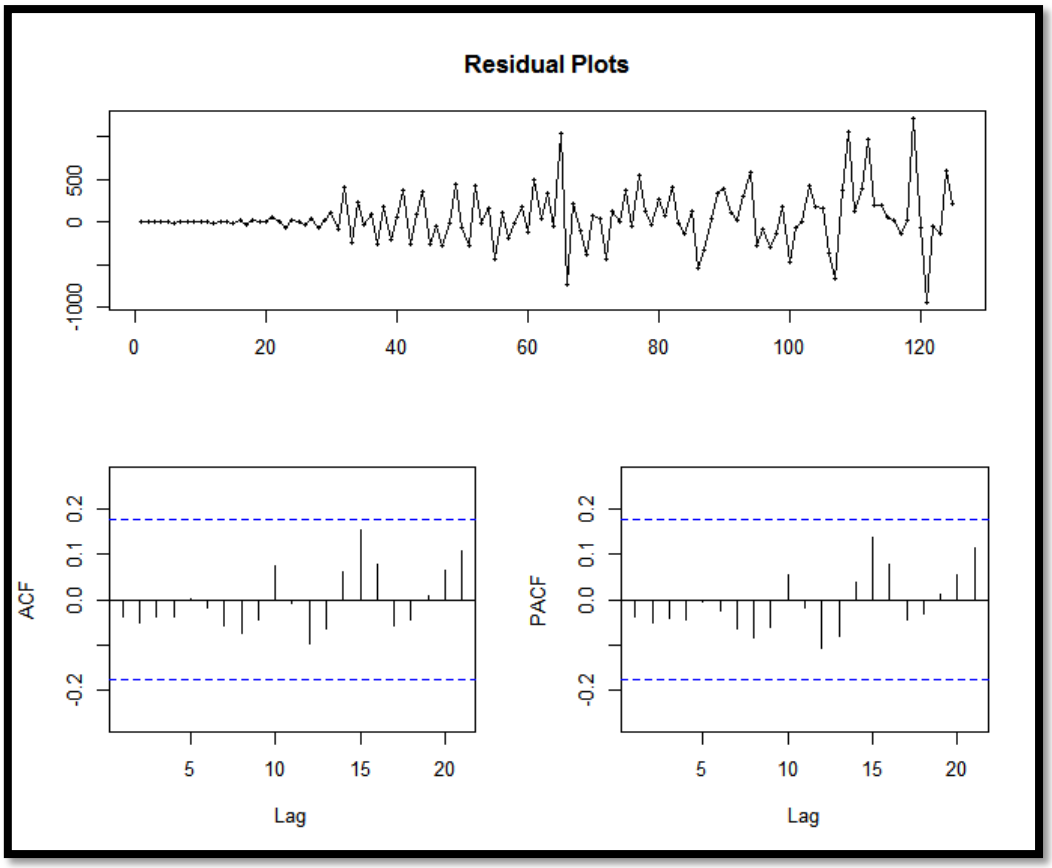


Figure 94: Residual plots of ARIMA (7,1,2) fitted for daily confirmed cases in The USA

Table 40: Box-Ljung Test for Residuals of ARIMA (7,1,2) Model

Box-Ljung test	
Data: Residuals from ARIMA (7,1,2)	
X-squared	0.037855
df=1	1
P value	0.8457

Box-Ljung test for the independence of residuals gave a *p* value 0.8457. The residuals were white noise (Table 40).

The fitted model is;

$$\Delta x_{t(USA)} = C+0.6956\Delta x_{t-1} \quad -0.2044\Delta x_{t-2} \quad - \quad 0.2995 \quad \Delta x_{t-3} \quad + \quad 0.2667\Delta x_{t-4}-$$

$$0.0153\Delta x_{t-5}+0.1060\Delta x_{t-6} +0.2447\Delta x_{t-7} -1.3843\varepsilon_{t-1}-0.7011\varepsilon_{t-2}+\varepsilon_t$$

Forecasts for 15 days were made. It showed a decrease in the daily number of confirmed cases for The USA.

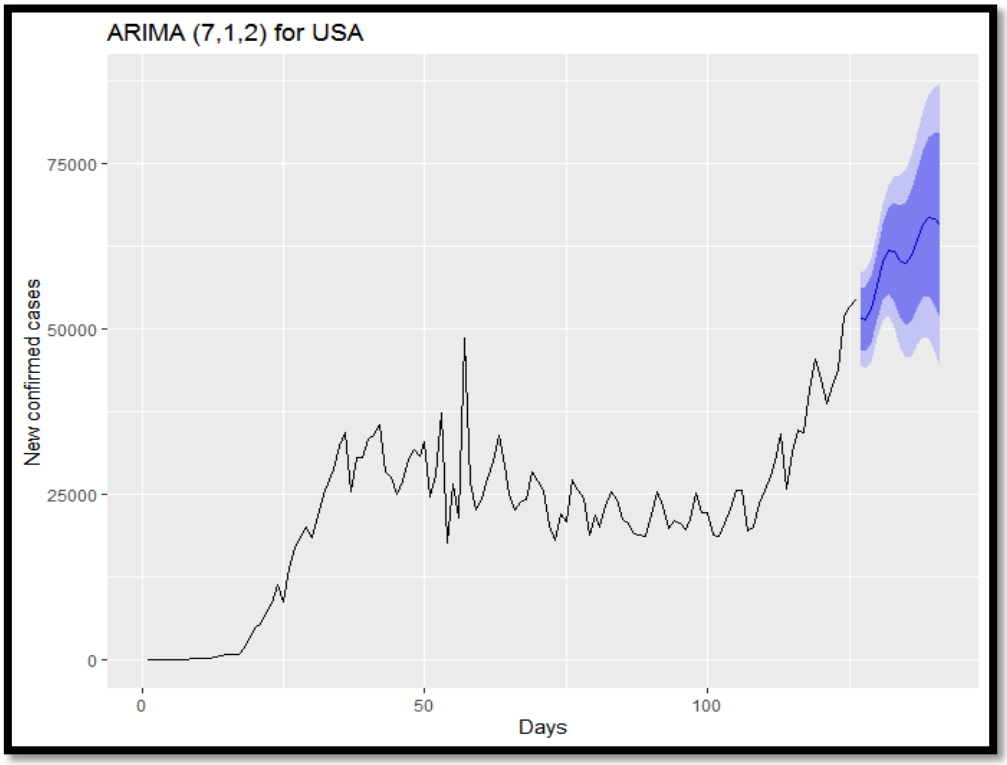


Figure 95: 15 days' forecasts for daily confirmed cases in The USA

Figure 95 gives the point and interval forecasts of daily counts of confirmed COVID-19 case in The USA for 15 days.

Italy

Time series of daily number of confirmed COVID-19 cases in Italy was plotted. The data was non-stationary. ACF and PACF plots showed significant spikes even beyond lag 10 (Figure 96). ADF test was performed to examine the stationarity of data.

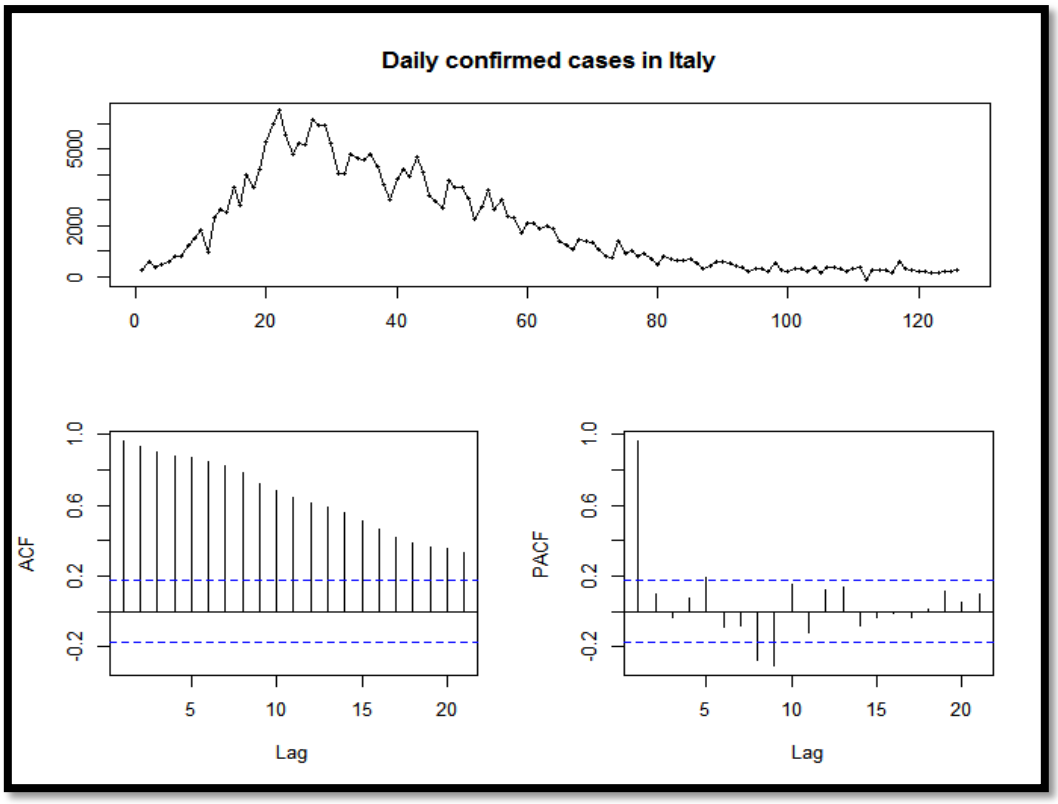


Figure 96: Time plots, ACF and PACF plots of daily confirmed cases in Italy

A non-significant *p* value suggested non-stationarity behaviour in the data. First order differencing was performed to eliminate any trend in the data (Table 41). Following figure 97 gives the plots done on differenced data.

Table 41: ADF test results for Italy

Augmented Dickey-Fuller test	
Data: Daily counts of Confirmed Cases in Italy	
Dickey-Fuller	-3.3143
Lag order	4
p-value	0.07205
Alternative Hypothesis: stationary	

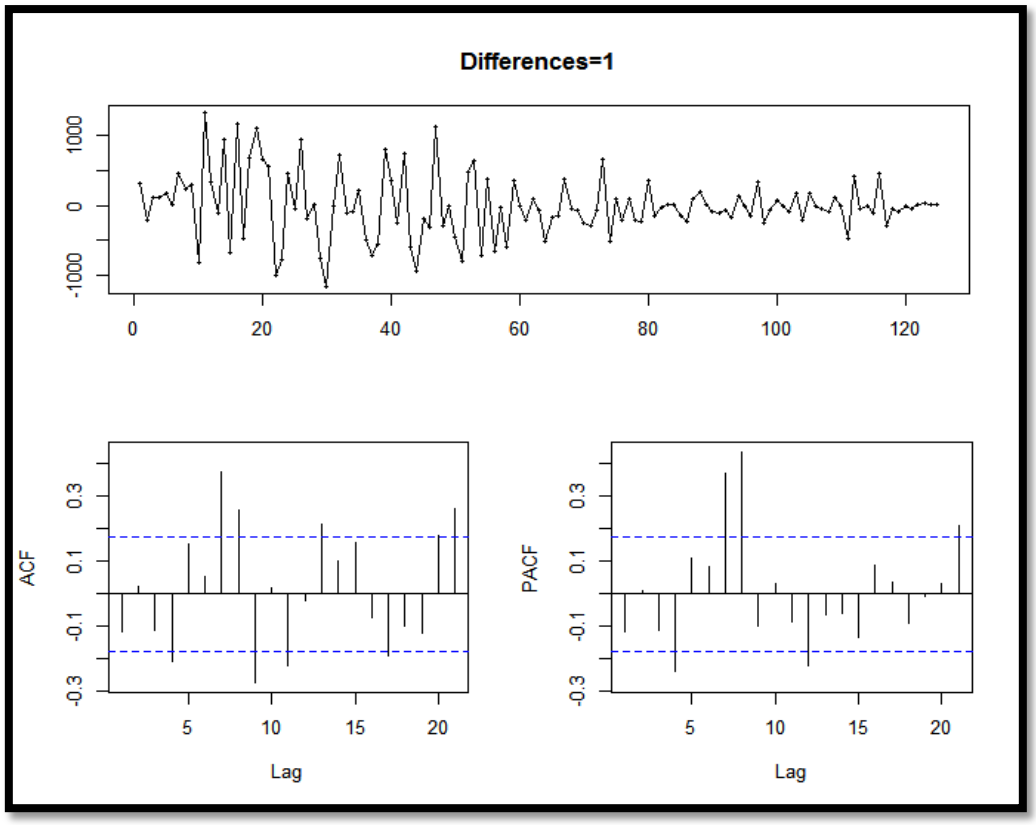


Figure 97: Time Plots, ACF and PACF plots of first order differenced data for confirmed COVID-19 cases in Italy.

ADF test was again performed. A significant *p* value = 0.01 was obtained. The alternative hypothesis that, data is stationary was thus accepted (Table 42).

Table 42: ADF Test for Stationarity of Daily Confirmed Cases in Italy

Augmented Dickey-Fuller test	
Data: Daily counts of Confirmed Cases in Italy	
Dickey-Fuller	-5.7962
Lag order	4
p-value	0.01
Alternative Hypothesis: stationary	

The lag 1 ACF showed a significant spike; indicating addition of more number of AR and MA terms. Several ARIMA models were developed to eliminate the pattern in ACF. The model with least AIC value was selected to be the best model. ARIMA (5,1,7) was the best model with least AIC value equals 1837.26. Table 43).

The competed models are given below;

Table 43: AIC Values for Competing ARIMA Models in Italy

Model (Italy)	AIC
ARIMA (4,1,7)	1846.46
ARIMA (5,1,7)	**1837.26**
ARIMA (5,1,8)	1841.23
ARIMA (6,1,7)	1839.26
ARIMA (7,1,4)	1844.01
ARIMA (8,1,3)	1847.31

Thus ARIMA (5,1,7) was fitted. In order to check if the fitted model was adequate, residuals were diagnosed for white noise. Residuals were plotted and Box-Ljung test was performed to examine the independence of residuals.

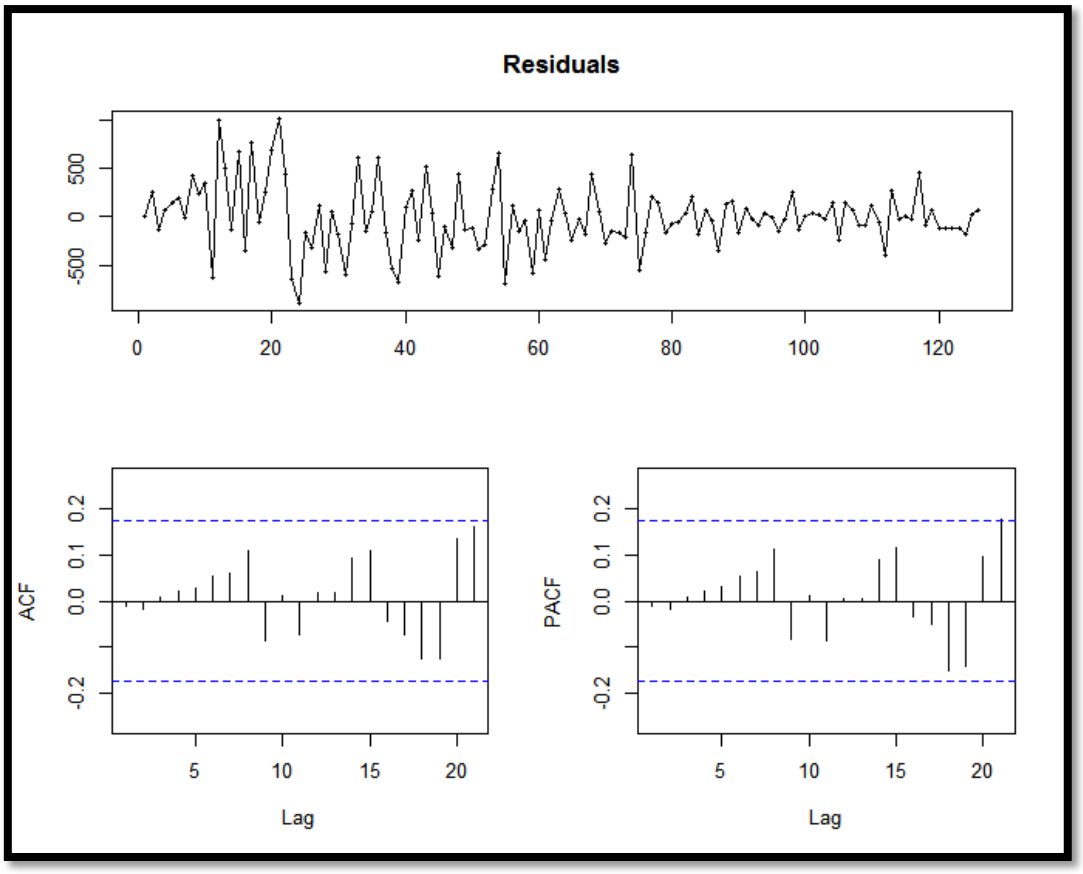


Figure 98: Residual plots of ARIMA (4,1,4) fitted for daily confirmed COVID-19 cases in Italy

Table 44: Box-Ljung Test for ARIMA (5,1,7) Residuals

Box-Ljung test	
Data: Residuals from ARIMA (5,1,7)	
X-squared	0.018447
df=1	1
P value	0.892

ACF and PACF plots did not show any significant spikes at any of the lag (Figure 98). Box-Ljung test resulted a non-significant *p* value concluding independence in residuals (Table 44).

Thus, model was proved to be adequate. The fitted model equation is;

$\Delta x_{t(Italy)} = C$+0.1659Δx_{t-1}-0.2931Δx_{t-2} +0.1289Δx_{t-3} - 0.4079Δx_{t-4} + 0.2441Δx_{t-5} -0.4766ε_{t-1}+ 0.3522ε_{t-2} − 0.4250ε_{t-3} + 0.5738ε_{t-4} -0.2148ε_{t-5} + 0.2025ε_{t-6} + 0.5026ε_{t-7}+ε_t

Point as well as interval (95% PI) forecasts of daily confirmed cases in Italy were made for 15 days. Table 45 and 46 present forecast results for COVID-19 cases with 95% prediction intervals across India, the USA, and Italy on various dates from July 5, 2020, to July 20, 2020. The lower and upper bounds of the prediction intervals are provided for each country, indicating the range of potential values. Table 47 provides point forecasts for the same set of dates, indicating the expected number of COVID-19 cases for India, the USA, and Italy. The following figure gives the pictorial representation of 15-day forecast (Figure 99).

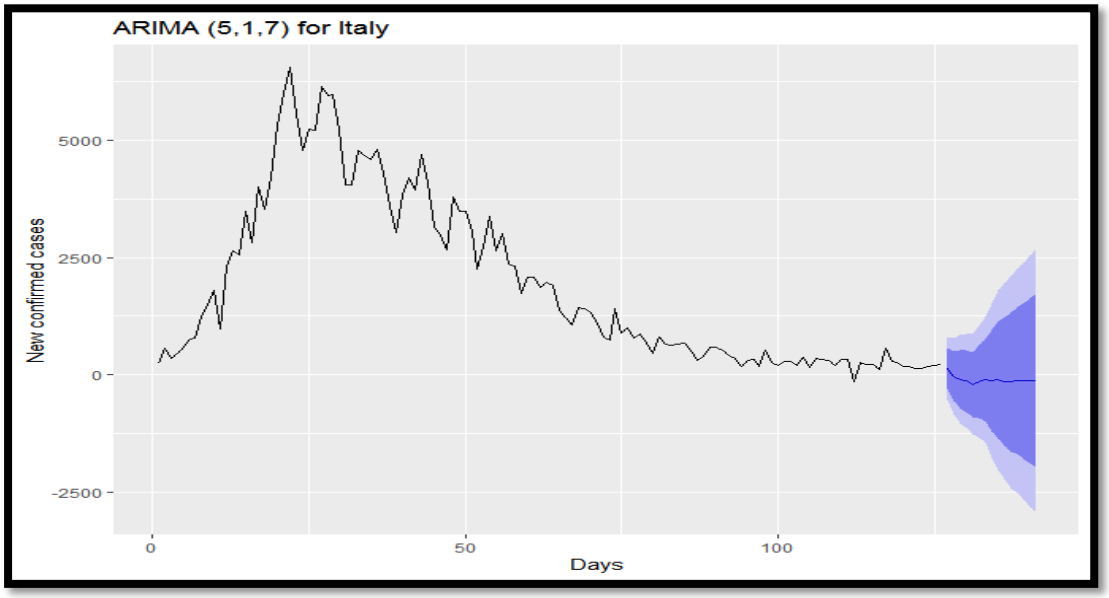


Table 45: First 10 Days' Prediction Intervals (95% CI) for COVID-19 Cases

Date	95% PREDICTION INTERVAL		
	India [Lower, Upper]	**USA** [Lower, Upper]	**Italy** [Lower, Upper]
05 July 2020	(22816.44, 24083.91)	(44400.70 , 58586.21)	(-523.71 , 805.77)
06 July 2020	(21655.30 ,23484.02)	(44036.44 , 58893.57)	(-840.24 , 778.55)
07 July 2020	(20466.59 , 22749.08)	(45136.60 , 60709.53)	(-1029.16 , 841.27)
08 July 2020	(20634.92 , 23192.12)	(48624.56 , 64411.45)	(-1128.68 , 862.46)
09 July 2020	(21615.63 , 24412.58)	(51264.20 , 68874.05)	(-1275.97 , 860.57)
10 July 2020	23085.57 , 26070.60)	(51968.17 , 71724.81)	(-1342.67 , 1048.70)

Table 46: Second 10 Days' Prediction Intervals (95% CI) for COVID-19 Cases

11 July 2020	(24278.55 , 27813.22)	(50297.78 , 72927.20)	(-1446.68 , 1236.90)
12 July 2020	(24111.52 , 28380.70)	(47293.56 , 73176.07)	(-1774.99 , 1525.49)
14 July 2020	(23548.35 , 28345.53)	(45585.62 , 74077.82)	(-2024.82 , 1801.44)
15 July 2020	(22651.02 , 27824.69)	(45861.48 , 76259.71)	(-2224.79 , 1942.27)
16 July 2020	(22359.02 , 27885.75)	(47586.79 , 79567.17)	(-2405.92 , 2096.13)
17 July 2020	(23195.23 , 28982.92)	(48861.70 , 82919.44)	(-2508.58 , 2251.71)
18 July 2020	(24585.54 , 30695.94)	(48570.43 , 85312.18)	(-265047 , 2382.42)
19 July 2020	(25573.27 , 32195.15)	(46658.96 , 86493.72)	(-2805.37 , 2546.92)
20 July 2020	(25821.72 , 32991.38)	(44278.89 , 87095.68)	(-2935.20 , 2685.14)

Table 47: Point Forecasts for COVID-19 Cases in India, the USA, and Italy

Date	Point forecast		
	India	**USA**	**Italy**
05 July 2020	23450.17	51493.45	141.03
06 July 2020	22569.66	51465.00	-30.84
07 July 2020	21607.84	52923.06	-93.95
08 July 2020	21913.52	56518.00	-133.11
09 July 2020	23014.10	60069.12	-207.70
10 July 2020	24578.10	61846.49	-146.98
11 July 2020	26045.89	61612.49	-104.89
12 July 2020	26246.11	60234.81	-124.74
14 July 2020	25946.94	59831.72	-111.69
15 July 2020	25237.86	61060.60	-141.26
16 July 2020	25122.38	63576.98	-154.89
17 July 2020	26089.07	65890.57	-128.43
18 July 2020	27640.74	66941.30	-134.03
19 July 2020	28884.21	66576.34	-129.22
20 July 2020	29406.55	65687.29	-125.03

- **Deaths, new cases and daily testing.**

The following table gives the daily counts of death. Maximum deaths occurred on a day is 2612 in The USA. India has the minimum number of deaths on a single day. The average number of daily deaths per day is over a thousand for the USA (Table 48).

Table 48: Daily COVID-19 Death Statistics by Country

Country	Daily counts of deaths					
	Min	**1st Quartile**	**Median**	**Mean**	**3rd quartile**	**Max**
India	0	2	28	35.02	51	194
USA	0	9	1134	1033	1985	2612
Italy	0	171	420	386.8	601.5	919

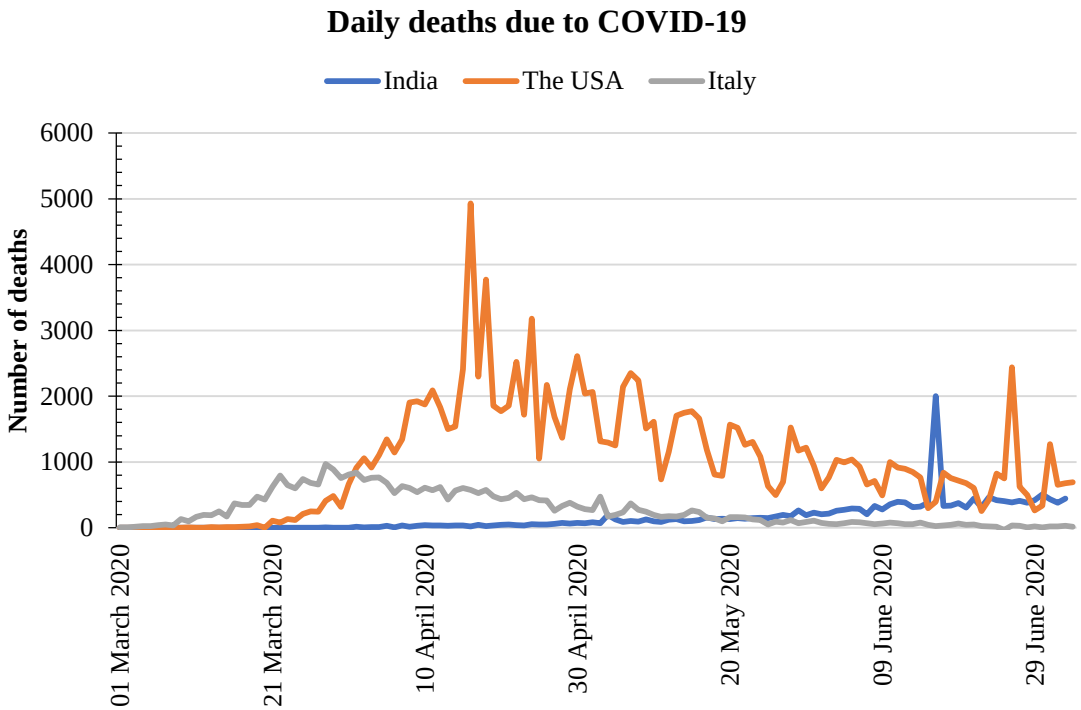


Figure 100: Daily counts of deaths in India, The USA and Italy.

Observing Figure of daily counts of deaths, The USA has a large number of daily deaths. India has the least. The USA leads in the maximum number of new cases reported in a single day. India has a maximum of 22771 cases (Figure 100).

Table 49 presents descriptive statistics of daily confirmed COVID-19 cases for India, the USA, and Italy. For India, the daily counts range from a minimum of 0 to a maximum of 22,771, with a median of 2,553 and a mean of 5,186. In the USA, the daily counts vary from 3 to 54,442, showcasing a median of 23,501 and a mean of 22,177. Italy, with daily counts ranging from 0 to 6,557, has a median of 1,079 and a mean of 1,907.

Table 49: Daily COVID-19 Confirmed Case Statistics by Country

Country	**Daily counts of confirmed cases**					
	Min	**1st Quartile**	**Median**	**Mean**	**3rd quartile**	**Max**
India	0	336	2553	5186	8909	22771
USA	3	18702	23501	22177	28386	54442
Italy	0	346	1079	1907	3316	6557

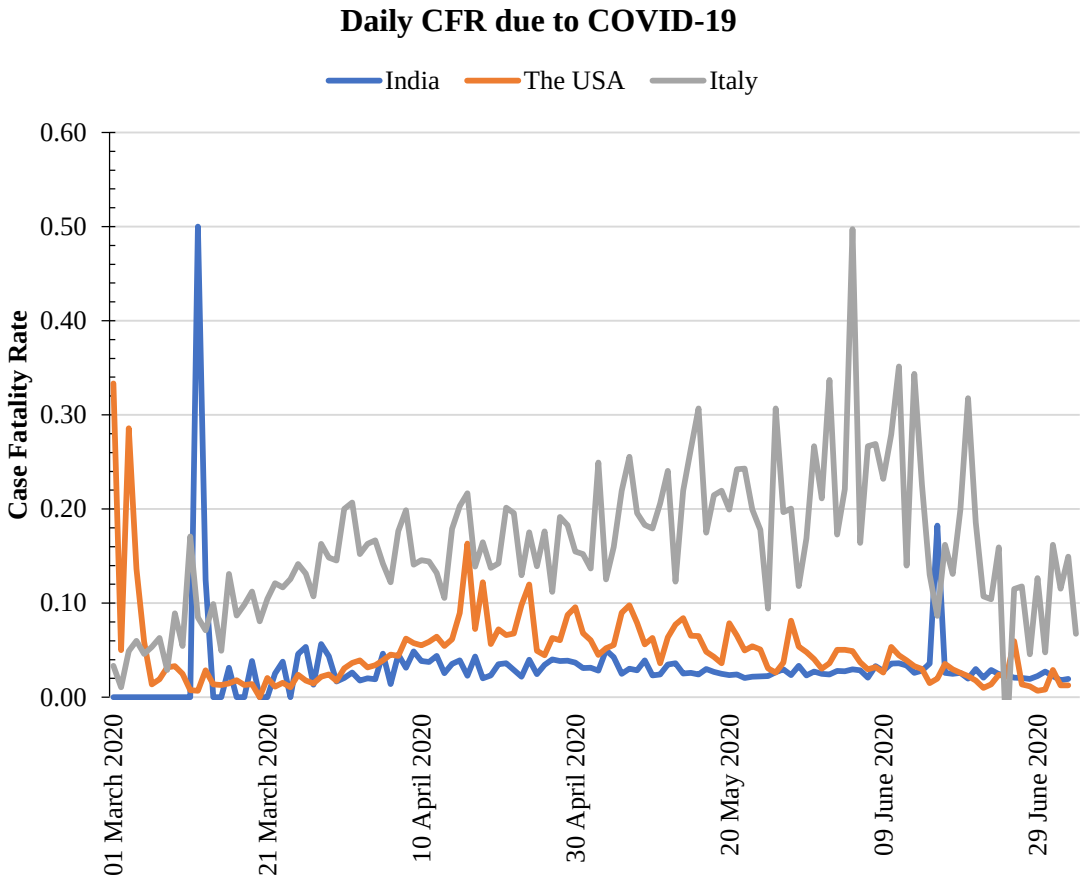


Figure 101: Daily case fatality rate of the three countries

The daily cases fatality rates for India was the lowest with mean 0.036 ± 0.04. Largest was for Italy with a mean of 0.126 ± 0.06. The mean case fatality rate for USA was 0.05 ± 0.02. At the beginning of the outbreak, India had a high CFR in the 3rd week of March (Figure 101). Total confirmed deaths for each country is given in Figure 102.

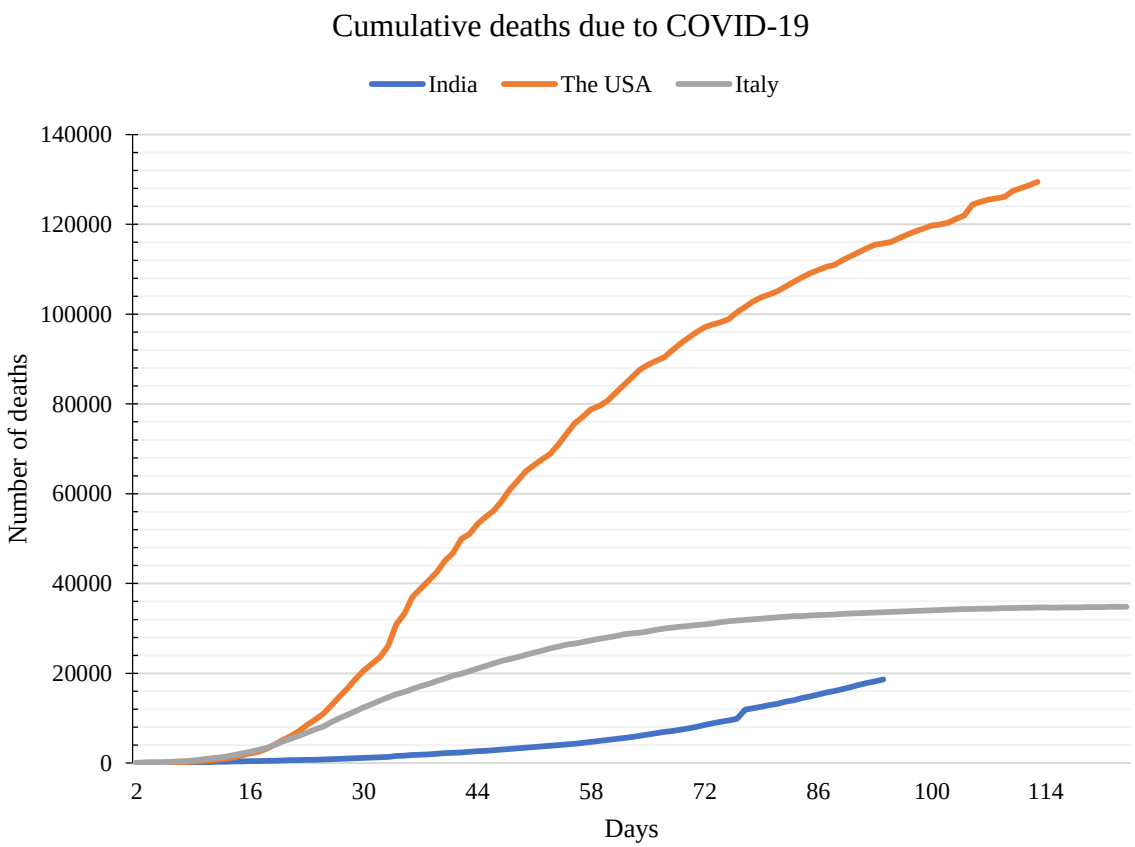


Figure 102: Total Confirmed COVID-19 deaths in India, The USA and Italy.

The figure shows 102 how rapidly deaths have increased for each country. The figure 102 shows how rapidly deaths have increased for each country. In India, increase in death count is not as fast as Italy and The USA.

The USA has the most number of PCR tests performed daily. Italy has the least. But accounting for their different population sizes, daily number of PCR tests conducted per 1000 people is given in Figure 103.

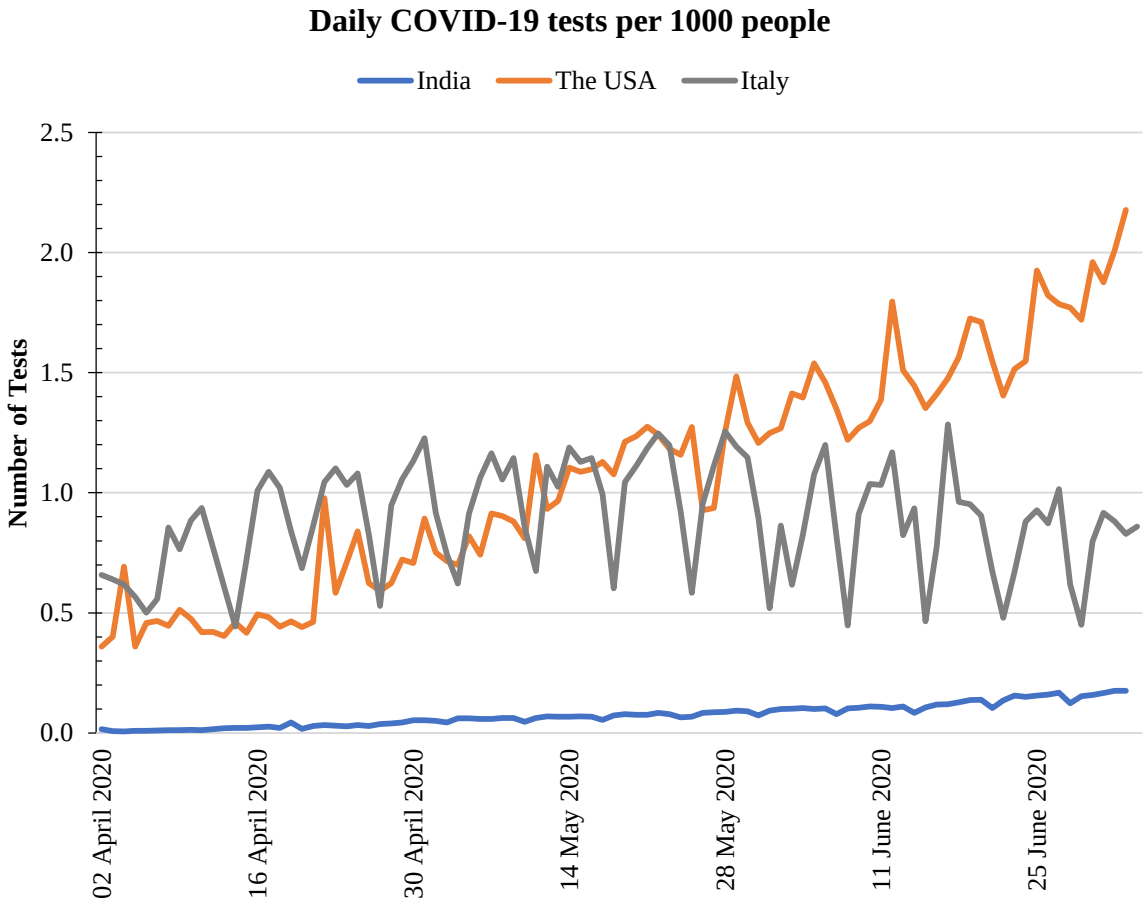


Figure 103: Daily number of COVID-19 tests performed per 1000 people (adjusted for population size).

Adjusted for population sizes, The USA and Italy have almost the same number of tests conducted per day. On 25 April 2020 Italy has 1.6 tests per 1000 people. On the same day, the USA has 0.6 tests per 1000 people. India is far away from the scene.

3.3 Making Decisions

3.3.1. Pollution Controllability

Table 50: Air quality ranges for AQIs.

AQI (CPCB 2015)	ORNAQI	$AQI_{Wei\ Av}$
0-50 Good	0 ≥ AQI ≤ 25 Clean	≥ AQI ≤ 0.5 Acceptable
51-100 Satisfactory	26 ≥ AQI ≤ 50 Light	0.51 ≥ AQI ≤ 1.0 Unacceptable
101-200 Moderately Polluted	51 ≥ AQI ≤ 75 Moderate	≥ AQI ≤ 2.0 Alert
201-300 Poor	76 ≥ AQI ≤ 100 Heavy	AQI ≥ 2.01 Significantly Harmful
301-400 Very Poor	≥ 101 Severe	
>401 Severe		

Table 50 outlines the Air Quality Index (AQI) ranges based on the Central Pollution Control Board's (CPCB) 2015 classification, the Outdoor Recreational Ambient Air Quality Index (ORNAQI), and the weighted average AQI (AQIWei Av).

Data obtained from CPCB is used to calculate average concentrations of ambient air pollutants PM_{10}, $PM_{2.5}$, SO_2 and NO_2 and the corresponding air pollution index. Different AQIs were estimated for periods before lockdown and during the lockdown and varying results were observed ranging from alert to good for the selected cities. This was due to the significant effect of lockdown in shrinking the levels of ambient air pollutants.

Variation in the concentration of individual air pollutants (PM_{10}, $PM_{2.5}$, SO_2 and NO_2) and air quality indices for pre and during lockdown periods are given in Table 51. The graphical representation of the table is given from Figure 104 to Figure 108 for the selected cities. The concentrations of pollutants got drastically reduced in all the cities during the shutdown period. These simultaneous variations in the concentration are depicting a lower pollution level in the selected areas. The combustion of fossil fuels in transport and industries are the major sources of these chemical pollutants. Thus, lockdown and related shutdown of industrial and transportation activities make the ambient air fresher and cleaner.

Table 51: Variation in the concentrations of pollutants and AQI.

City	Time Period	Pollutants				AQI			
		PM_{10}	$PM_{2.5}$	NO_2	SO_2	AQI_{mean}	AQI_{gm}	ORNAQI	AQI_{WeiAv}
Ghaziabad	BL	196.10	108.48	53.78	17.94	115.43	82.22	174.45	1.390
	DL	101.82	41.26	20.49	17.54	52.94	43.29	82.07	0.634
Delhi	BL	178.80	107.70	41.45	15.90	99.08	56.72	130.52	1.279
	DL	124.82	49.57	27.74	13.81	60.30	40.23	80.85	0.763
Kolkata	BL	181.201	96.17	36.50	09.07	177.73	60.61	130.88	1.238
	DL	59.36	25.38	06.12	04.56	45.39	17.39	39.49	0.375
Hyderabad	BL	99.33	44.05	27.91	09.52	54.88	40.55	73.87	0.648
	DL	62.04	28.68	16.92	08.41	35.40	27.72	48.38	0.417
Cochin	BL	102.98	16.51	22.15	08.74	42.28	26.91	57.29	0.575
	DL	58.01	09.82	06.16	05.53	22.25	14.64	30.88	0.307

Note: DL- During Lockdown, BL- Before Lockdown

The variations in air quality indices calculated by four different methods for all the five cities due to the lockdown in India are shown in Table and Figures. It was noted that the higher value of an index refers to the greater level of air pollution and consequently greater health issues. From Table 51, it is clear that the concentrations of gaseous air pollutants have diminished due to the restrictions on human and industrial activities. Therefore, air quality indices also diminished correspondingly. By comparing AQIs calculated based on four methods with predefined quality ranges provide substantial evidence for this improvement in the quality of air. When pollution rates are considered, before the lockdown begins, Ghaziabad and Kolkata were the most polluted cities in the list. But when the lockdown begins, both the cities make sudden changes in the air quality compared to other cities. Among five of the selected cities, Cochin remains the least polluted in both the periods. It was concluded that the ambient air quality of Indian cities is improved with acceptable levels of deviations in the index values.

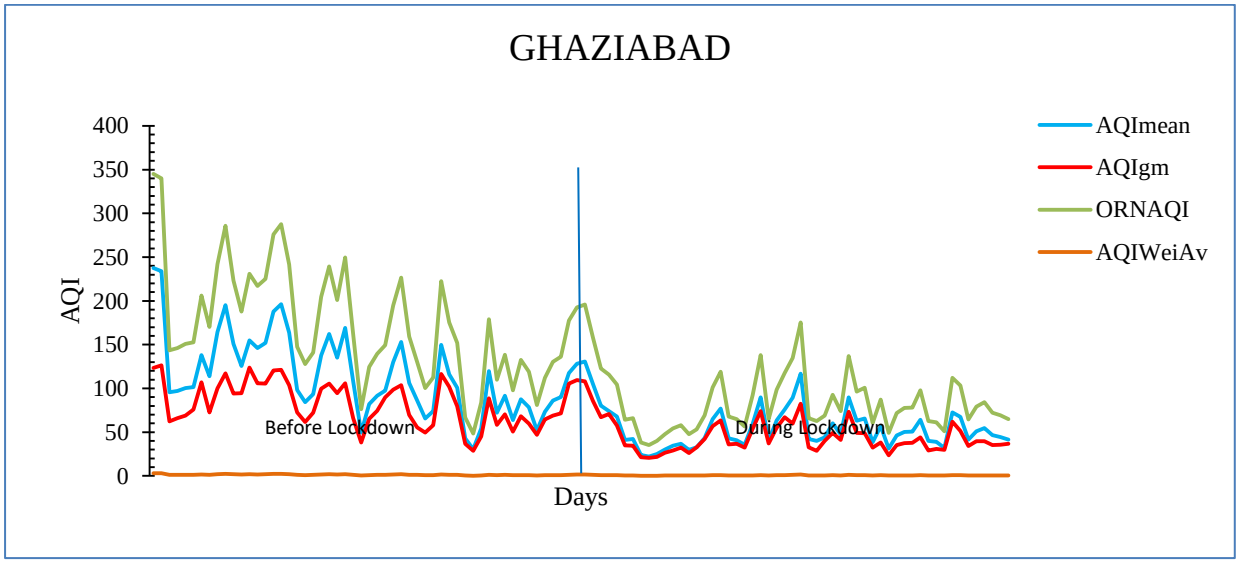


Figure 104: Variation of air quality indices at Ghaziabad

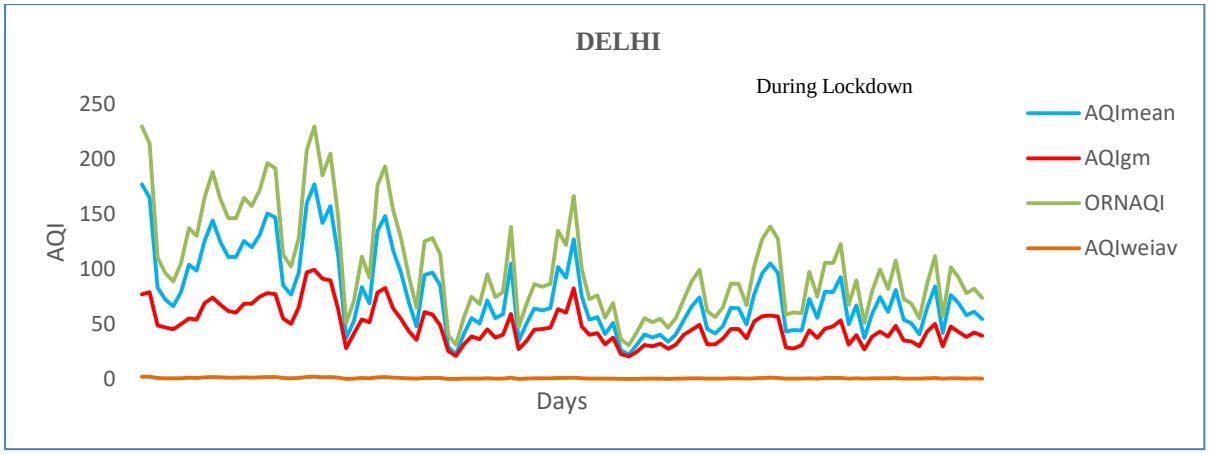


Figure 105: Variation of air quality indices at Delhi

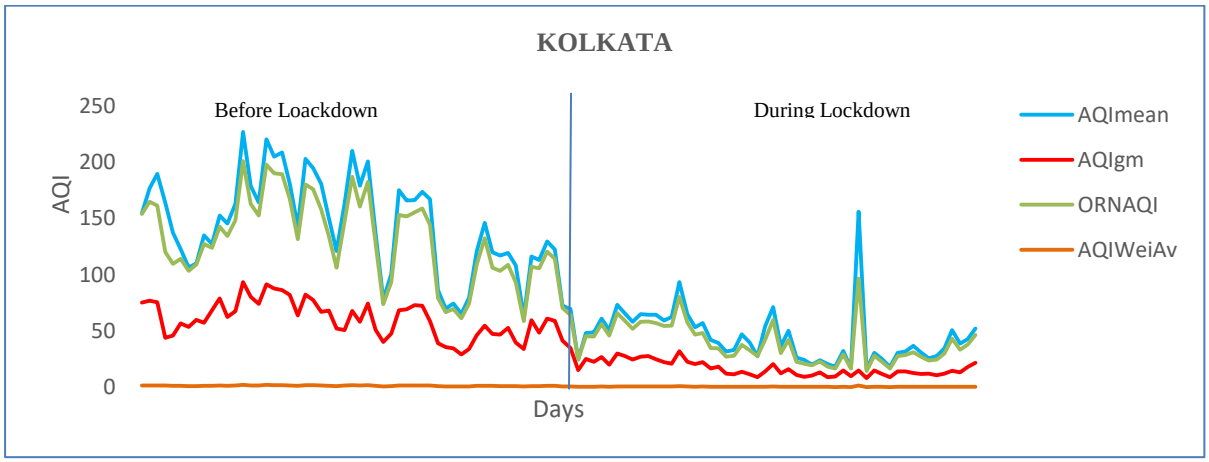


Figure 106: Variation of air quality indices at Kolkata.

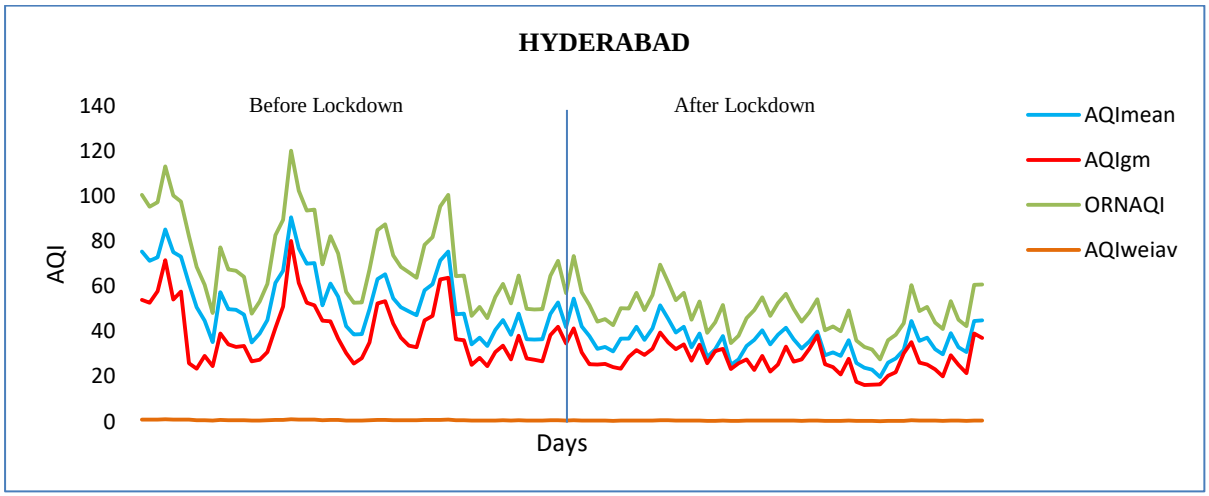


Figure 107: Variation of air quality indices at Hyderabad

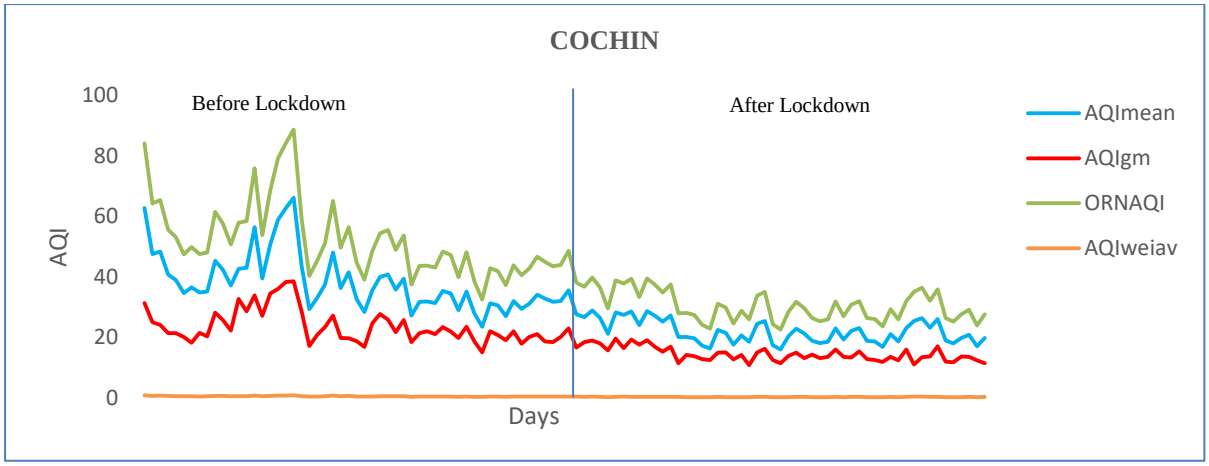


Figure 108: Variation of air quality indices at Cochin.

The Paired t-test is used to test the significant difference between two time periods (pre and during lockdown). Here the null hypothesis, the true mean difference between the paired samples is zero, is tested against the alternative hypothesis that the true mean difference of the paired samples is not equal to zero. Table records the significant difference in the mean concentrations of ambient air pollutants PM_{10}, $PM_{2.5}$, SO_2 and NO_2 at these cities during the comparative period.

The difference in the mean values of all parameters in respect of two time periods (before and after lockdown) is statistically significant ($p > 0.05$) except the difference in the concentration of SO_2 at Ghaziabad. It was said that there is a significant difference between the mean concentration of pollutants before and during the lockdown (Table 52).

Table 52: Mean difference and its significance using t test

City	Pollutants											
	PM_{10}			$PM_{2.5}$			NO_2			SO_2		
	MD	95% CI	p	MD	95% CI	p	MD	95% CI	p	MD	95% CI	p
Gh	94.2	(69.9, 118.6)	<0.001	67.5	(50.0, 82.2)	<0.001	33.2	(27.1, 37.2)	<0.001	00.3	(-3.9, 2.7)	0.739
Del	48.9	(24.3, 73.6)	0.0001	51.0	(36.7, 65.3)	<0.001	15.8	(12.3, 19.4)	<0.001	03.6	(-2.8, 0.1)	0.081
Kol	121.6	(104.1, 139.1)	<0.001	70.7	(61.1, 180.4)	<0.001	30.3	(28.1, 35.6)	<0.001	4.5	(3.3, 5.6)	<0.001
Hyd	37.5	(29.7, 45.2)	<0.001	13.1	(9.1, 17.1)	<0.001	14.5	(8.4, 13.1)	<0.001	01.3	(0.8, 3.5)	0.0233
Coc	38.2	(32.3, 44.2)	<0.001	06.7	(2.1, 11.3)	0.004	10.9	(7.7, 14.2)	<0.001	00.6	(0.3, 0.9)	0.003

Note: MD- Mean Difference, p- Pvalue, Gh- Ghaziabad, Kolkata- Kol, Hyderabad- Hyd, Delhi- Del, Cochin Coc

Correlation coefficients obtained by the Pearson coefficient of correlation for the cities during the study periods are shown in Figure 109 to Figure 113. This shows that before lockdown, the cities have a positive correlation between the parameters PM_{10}, $PM_{2.5}$, SO_2 and NO_2 except at Cochin. During the pre-lockdown period Ghaziabad, Delhi and Kolkata showed a higher correlation between Particulate Matters 10 and 2.5. Also, for Ghaziabad and Delhi there exists the least correlation between SO_2 and PM_{10} during both the periods. But in Delhi during the lockdown period, the higher correlation remains the same but the least correlation is between the pollutants NO_2 and PM_{10}. It should be noted that de deviation in the SO_2 levels of Ghaziabad is not significant. While in Hyderabad there is a higher correlation between PM_{10} and SO_2

and a lower correlation between NO_2 and $PM_{2.5}$ before the lockdown and during lockdown there exists a higher correlation between NO_2 and SO_2 and a lower correlation between PM_{10} and NO_2. But in the case of Cochin, the correlation between the pollutants are very small and the pollutants NO_2 and PM_{10} show a negative correlation at the pre lockdown period. During the lockdown, the negative correlation exists between SO_2 and PM_{10}. In brief, the strength of the correlation between the pollutants is reduced during the lockdown period.

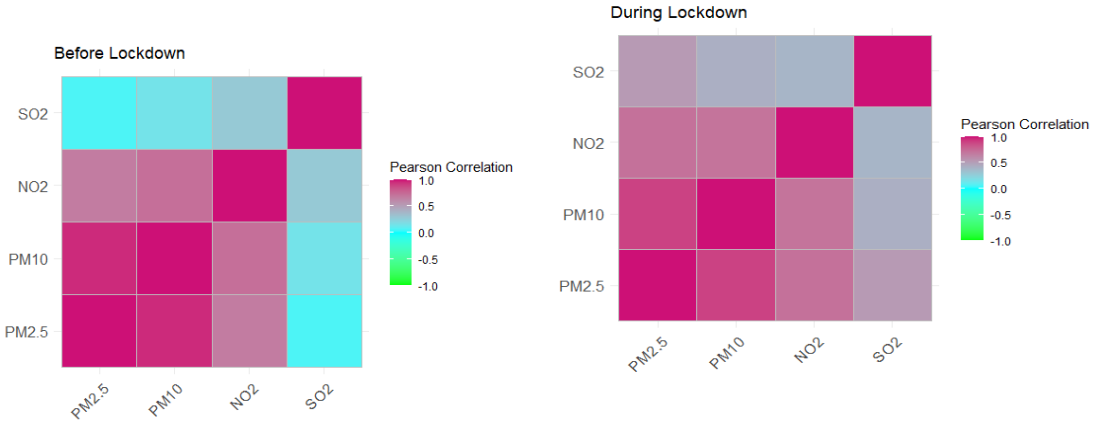


Figure 109: Correlation between pollutants before and during the lockdown.

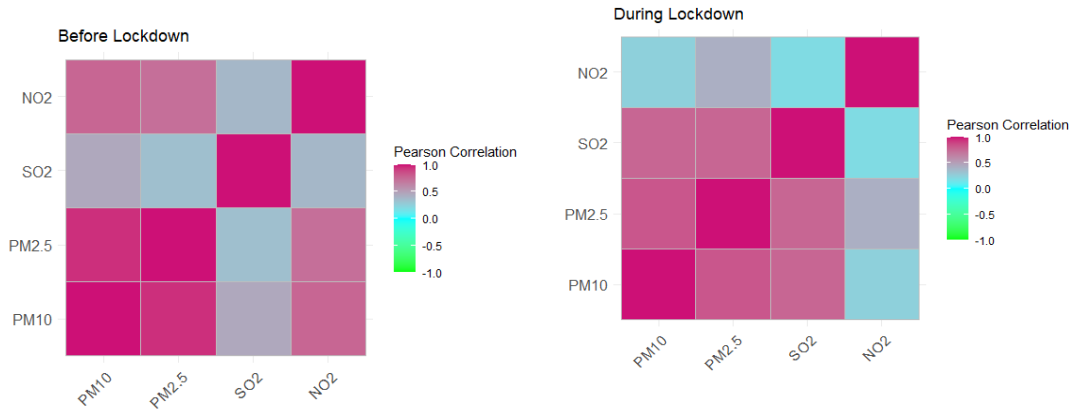


Figure 110: Correlation between pollutants at Ghaziabad before and during the lockdown.

Figure 111: Correlation between pollutants at Delhi before and during the lockdown.

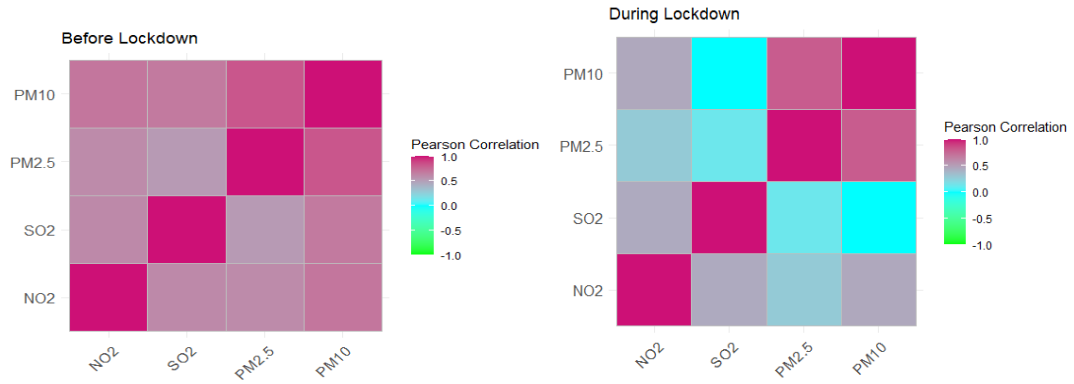


Figure 112: Correlation between pollutants at Kolkata before and during the lockdown.

Figure 113: Correlation between pollutants at Hyderabad before and during the lockdown.

With the effect of lockdown, all the four pollutants PM_{10}, $PM_{2.5}$, SO_2 and NO_2 undergo a decreasing trend in the concentrations at the selected cities (Figure 114 to Figure 118). It is shown that average PM_{10} levels are reduced by 61%, 30%, 68%, 37%, and 43% at Ghaziabad, Delhi, Kolkata, Hyderabad and Cochin respectively. Similarly, in the case of $PM_{2.5}$, the average concentrations are reduced at a rate of 61%, 53%, 73%, 34%, and 40% respectively. At Ghaziabad, Delhi, Kolkata, Hyderabad and Cochin, NO_2 is reduced at 61%, 33%, 83%, 39%, and 72%. The decrease in the concentrations of SO_2 is very low when compared to other pollutants. SO_2 undergoes a decline of 2% at Ghaziabad, 13% in Delhi, 49% at Kolkata, 11% at Hyderabad and 36% at Cochin. It is estimated that due to the rapid decrease in vehicular emissions, industrial and construction shutdowns, concentrations of all the pollutants get diminished.

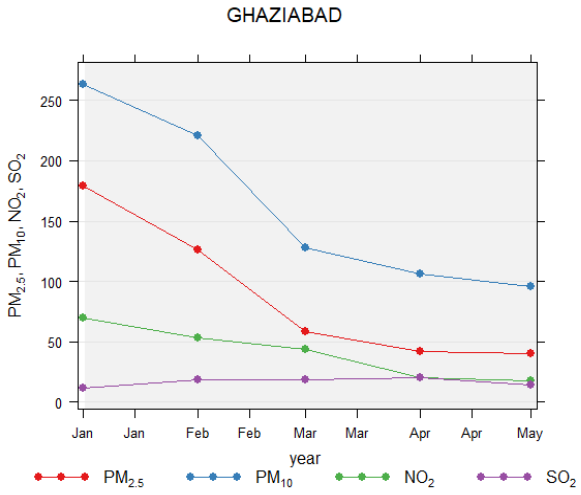


Figure 114: Trend in the variation of pollutant concentration at Ghaziabad.

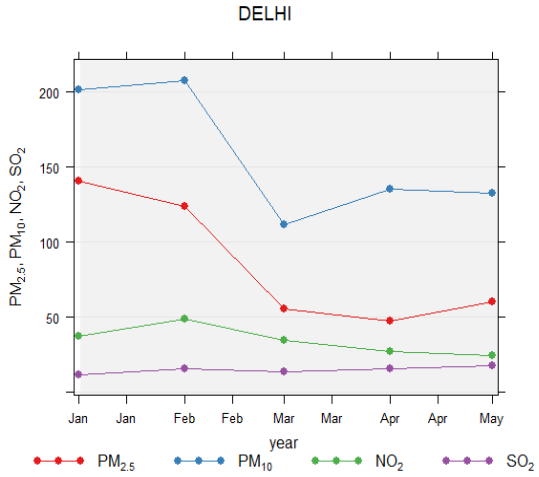


Figure 115: Trend in the variation of pollutant concentration at Delhi.

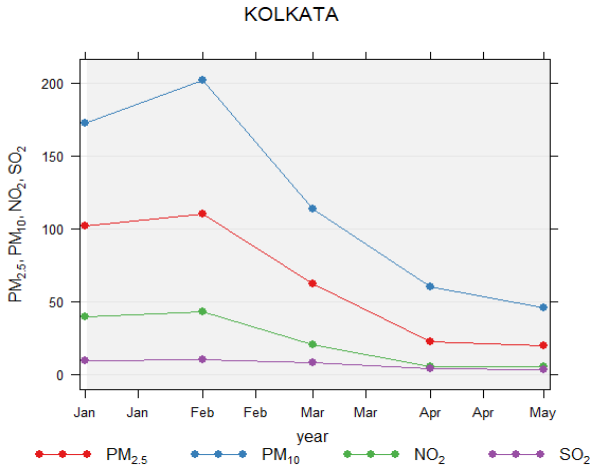


Figure 116: Trend in the variation of pollutant concentration at Kolkata.

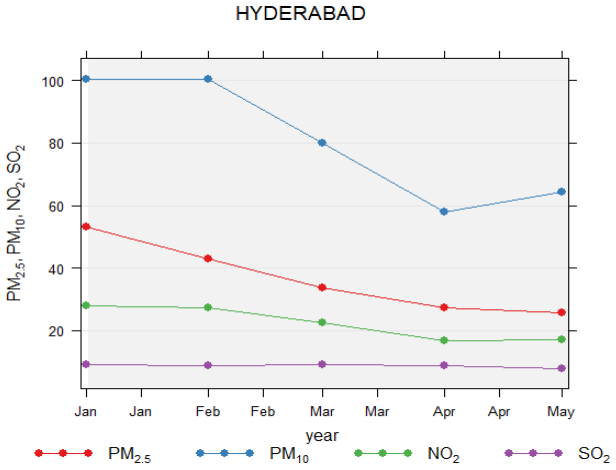


Figure 117: Trend in the variation of pollutant concentration at Hyderabad.

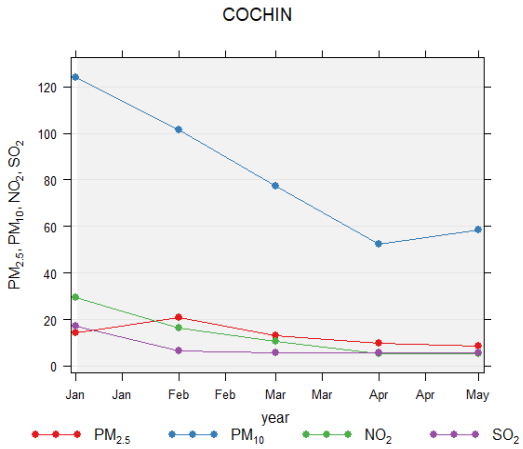


Figure 118: Trend in the variation of pollutant concentration at Cochin

The calendar plots (Figure 119 to Figure 123) are used to represent the concentrations of pollutants during the study period in a conventional calendar format i.e. by month and day of the week. The calendar heat map visualizes the daily patterns in the variation of the pollutants PM_{10}, $PM_{2.5}$, SO_2 and NO_2 among the selected cities. The plots given in Figure clearly emphasizes that the concentrations of pollutants were at the peak during the months January, February and March (before the lockdown period). It refers to the higher air pollution experienced by the populations living in urban areas due to excessive emission of exhaust fumes from vehicles, power generation and manufacturing processes. From March 20 onwards, the pollutants represent a progressive decline in their average concentrations. Since air traffic is down, factories are closed and fewer vehicles are on the road, the ambient air is positively thriving during the lockdown period.

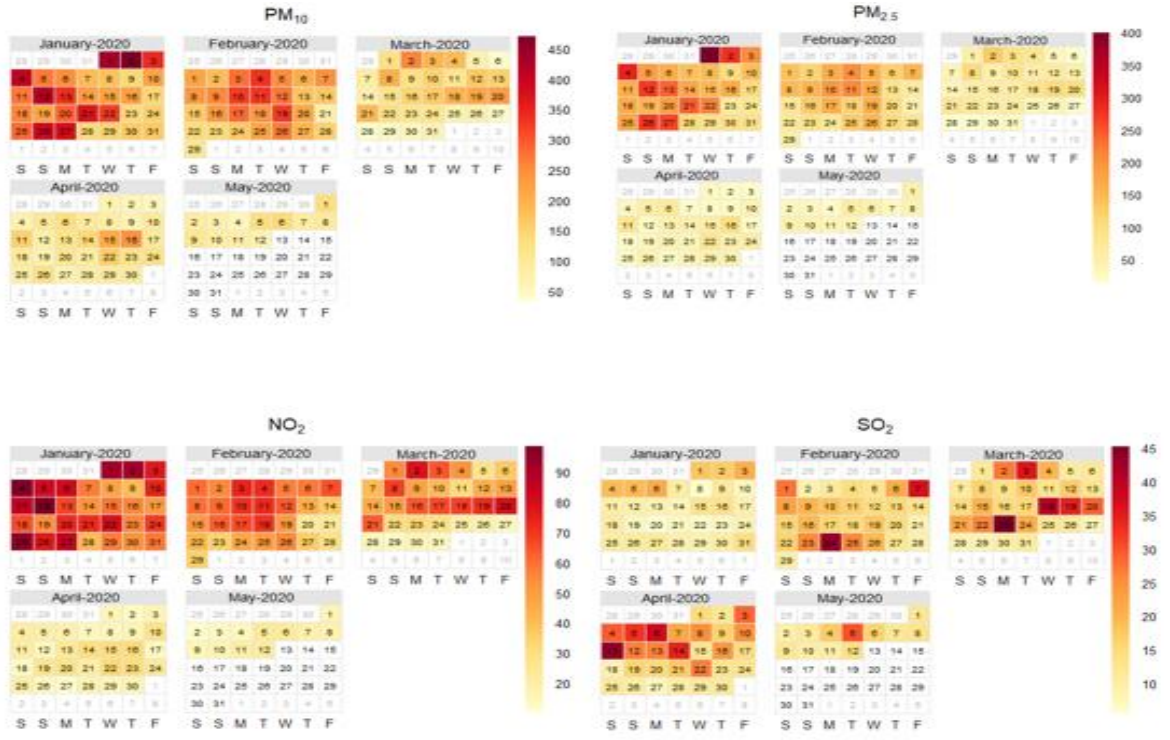


Figure 119: Calendar plot for the concentration of pollutants PM_{10}, $PM_{2.5}$, SO_2 and NO_2.

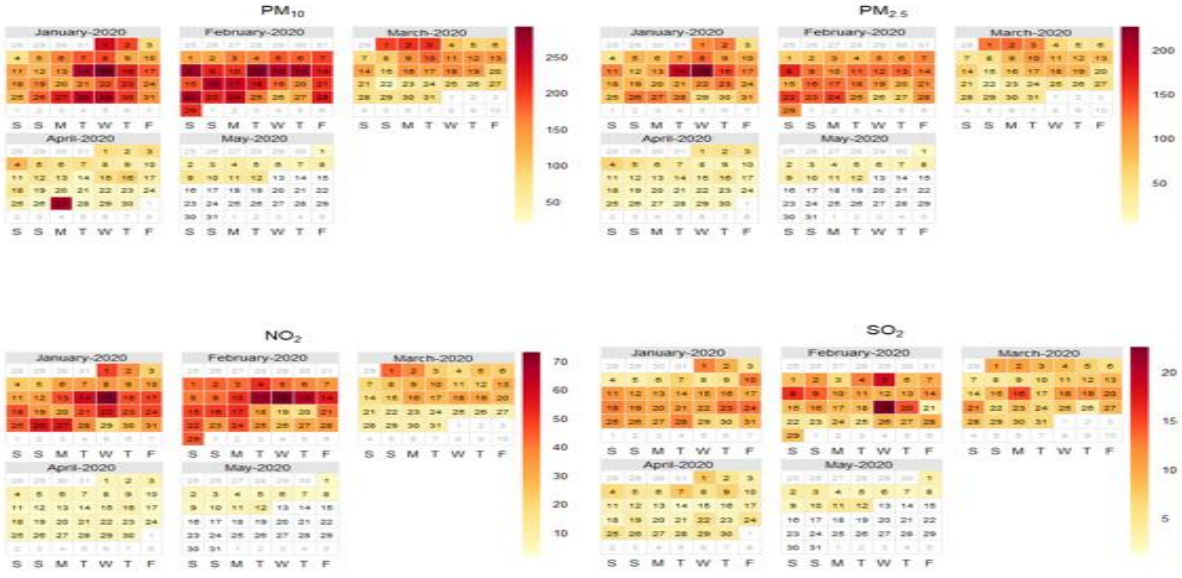


Figure 120: Calendar plot for concentration of pollutants at Ghaziabad.

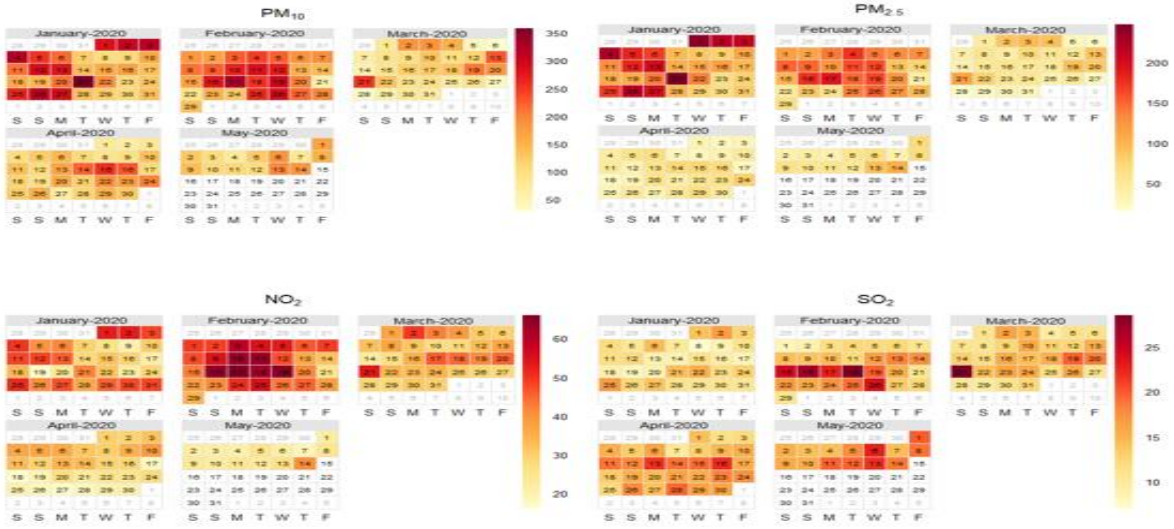


Figure 121: Calendar plot for concentration of pollutants at Delhi.

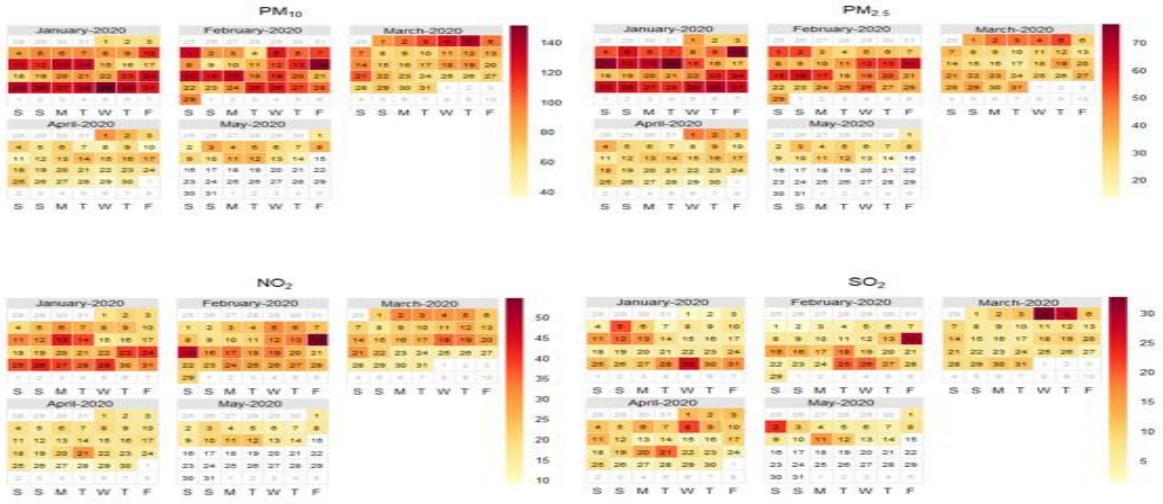


Figure 122: Calendar plot for concentration of pollutants at Kolkata.

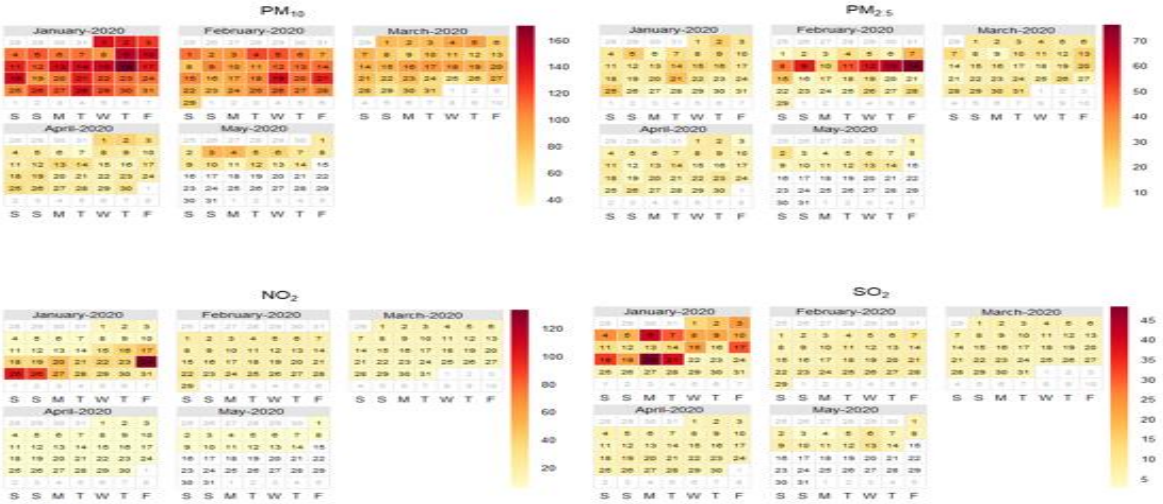


Figure 123: Calendar plot for concentration of pollutants at Hyderabad.

Time plot is a technique for visualizing the data with some measure of time is the unit taken on the horizontal axis (Figure 124 - Figure 128). Here the graph is generated by taking the duration of months January to May 2020. Figure gives the practical exhibition of the changes in the concentrations of pollutants PM_{10}, $PM_{2.5}$, SO_2 and NO_2 over this time for the selected cities. Among all these cities, pollutants were showing higher ranges of concentrations during the pre-lockdown period and are fluctuated randomly with some extreme measurements or outliers of average concentrations. From the end of March, Indian cities are under lockdown as a response to COVID-19 pandemic. The pollutants have noticeably reduced and approximately remain constant with lower values with fewer fluctuations in concentrations.

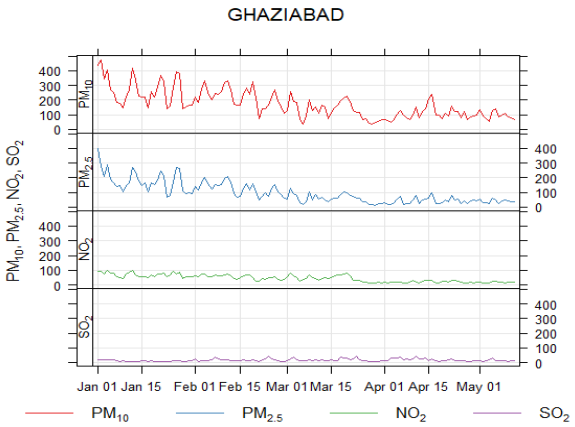


Figure 124: Time plot for pollutant concentrations at Ghaziabad.

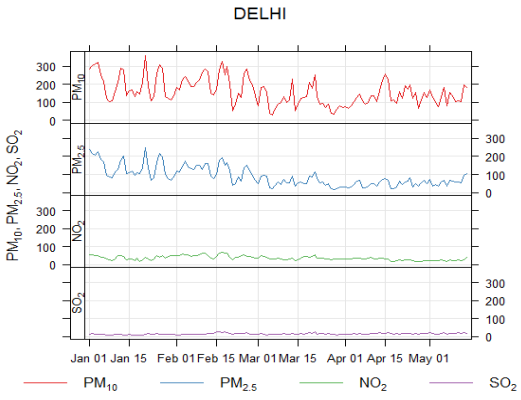


Figure 125: Time plot for pollutant concentrations at Delhi.

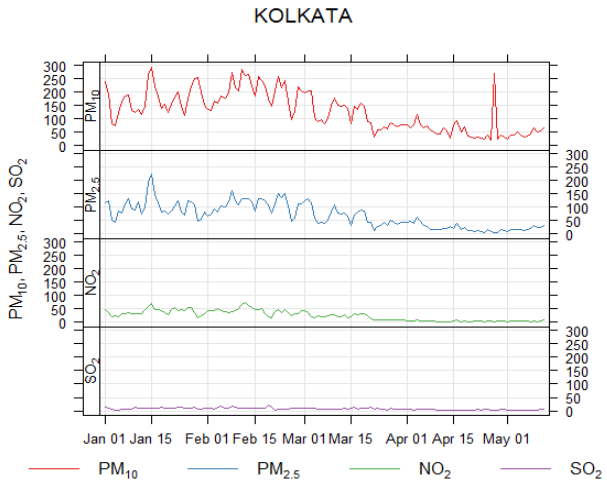


Figure 126: Time plot for pollutant concentrations at Kolkata.

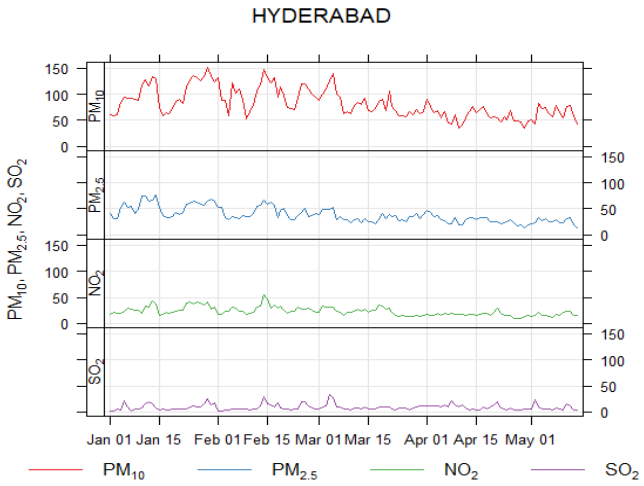


Figure 127: Time plot for pollutant concentrations at Hyderabad.

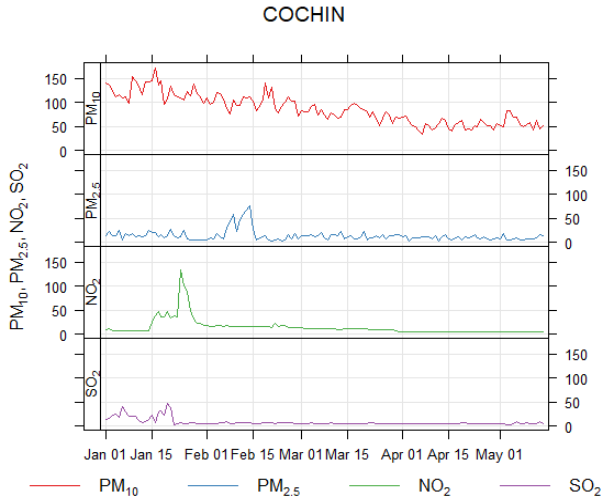


Figure 128: Time plot for pollutant concentrations at Cochin.

All the orders of p, d and q and chosen a model which have least AIC value. Also, the selected model adequately meets all the expectations of residual analysis.

Dataset consisting of pollutants PM_{10}, $PM_{2.5}$, SO_2 and NO_2 into a time series one and check for stationary of data separately for each pollutant of selected cities using Augmented Dickey Fuller Test. The null hypothesis the data is not stationary is tested against alternative hypothesis that the data is stationary. The level of significance (LOF) is taken as 0.05. If p value exceeds LOF, accept the null hypothesis and if the p value lies within the LOF, reject the null hypothesis that the data is not stationary. The respective p values are given in Table 53 for each pollutant of the respective cities. If data is not found to be stationary useing differencing to make the mean of the time series which will make the time series data stationary by eliminating trend and seasonality. If single differencing do not attain stationarity, it was repeated until the data becomes stationary.

Table 53: Results of ADF Test

Cities	**Respective p-values**			
	PM_{10}	**$PM_{2.5}$**	**NO_2**	**SO_2**
Ghaziabad	0.038	0.320	0.615	0.012
Delhi	0.123	0.683	0.331	0.045
Kolkata	0.103	0.043	0.068	0.010
Hyderabad	0.590	0.010	0.010	0.014
Cochin	0.069	0.069	0.084	0.029

The Kwiatkowski- Philips- Schmidt- Shin (KPSS) test is a type of unit root test which determines the stationarity of time series data around a deterministic trend. KPSS test is based on Large Multiplier (LM) test with null hypothesis that the random walk has a zero variance against the alternative that random walk has a non-zero variance. The $p < 0.05$ was considered as the significance level. Thus, the hypothesis H_0 of stationarity will be rejected if the observed value is less than 0.05. The respective p values of the pollutants of each city is given in Table 54.

Table 54: Results of KPSS Test

Cities	Respective p-values			
	PM_{10}	**$PM_{2.5}$**	**NO_2**	**SO_2**
Ghaziabad	0.100	0.010	0.032	0.100
Delhi	0.010	0.010	0.010	0.100
Kolkata	0.010	0.014	0.010	0.018
Hyderabad	0.010	0.100	0.060	0.100
Cochin	0.069	0.069	0.100	0.010

Plot ACF and PACF for each pollutant of these cities to identify the ARIMA model. Then was obtained the final model for original data AIC values. The AIC essentially chooses a model with the best fit, as measured by the likelihood function, subject to a penalty term that increases with the number of parameters fitted in the model. The Table 55 shows the selected AIC value of each matrix and helps us to find the best values for *p* and *q* and corresponding ARIMA models.

Table 55: AIC values and selected ARIMA Models

Cities	Pollutants							
	PM_{10}		$PM_{2.5}$		NO_2		SO_2	
	AIC	Selected ARIMA	AIC	Selected ARIMA	AIC	Selected ARIMA	AIC	Selected ARIMA
Gh	1457.5	(1,1,1)	1438.5	(2,1,1)	1106.7	(2,1,2)	966.7	(2,1,1)
Del	1541.1	(2,1,2)	962.1	(2,1,1)	957.6	(2,1,2)	755.5	(2,1,1)
Kol	1464.1	(2,1,2)	1282.4	(0,1,3)	943.7	(3,1,2)	678.3	(1,1,3)
Hyd	927.3	(2,1,3)	740.9	(1,1,1)	658.8	(3,1,3)	714.6	(1,1,3)
Coc	937.5	(3,1,3)	937.5	(3,1,3)	267.0	(3,1,3)	541.4	(3,1,3)

Note: Gh- Ghaziabad, Kolkata- Kol, Hyderabad- Hyd, Delhi- Del, Cochin- Coc

Next, checking the adequacy of the model. To have a 'good' model, expect the residuals to be 'random' and 'close to zero' and mode validation is carried out by ACF plots. Clearly shows that there is no significant correlation between the residuals as all the points lie within the boundary lines. The Ljung Box test (Table 56) examines the autocorrelation of residuals and if the autocorrelation is very small it is clear that the model adequately fits the data.

The normality of errors is confirmed from QQ-plot of residuals and Shapiro-Wilk test (Table 57). For both the tests, $p < 0.05$ was considered as the significance level and the null hypothesis are rejected when the observed value lies below the LOF. The results are given in table 56.

Table 56: Results of Ljung Box test

Cities	Respective P-values			
	PM_{10}	**$PM_{2.5}$**	**NO_2**	**SO_2**
Gh	0.7597	0.7597	0.8384	0.7997
Del	0.8272	0.6026	0.9225	0.9122
Kol	0.5769	0.8722	0.6931	0.9578
Hyd	0.7207	0.9648	0.7185	0.2169
Coc	0.902	0.902	0.982	0.8267

Note: Gh- Ghaziabad, Kolkata- Kol, Hyderabad- Hyd, Delhi- Del, Cochin- Coc

Table 57: Shapiro-Wilk Normality Test

Cities	Respective P-values			
	PM_{10}	**$PM_{2.5}$**	**NO_2**	**SO_2**
Gh	0.039	0.039	0.2586	8.377e-06
Del	0.6031	0.3911	0.2345	0.0126
Kol	2.181e-09	0.0011	1.595e-06	7.713e-05
Hyd	0.0306	0.0472	0.0486	6.33e-10
Coc	2.341e-12	2.341e-12	6.299e-12	1.811e-12

Note: Gh- Ghaziabad, Kolkata- Kol, Hyderabad- Hyd, Delhi- Del, Cochin- Coc

Table 58 shows predicted value of the air pollutants 4 weeks after 12th May 2020. It indicates the details clearly that lockdown in India benefited a lot to reduce pollution. The AQI values in Table shows the quality of air increased a lot. A significant decrease in AQI on comparing to AQI before the lockdown was seen. All cities' air quality is good according to the quality scale provided by CPCB. Before the lockdown, all the cities except Kochi were in satisfactory range. If these major cities reduced to good and was assumed that lockdown made India into a good ambient condition.

Table 58: Forecasted Values of the Pollutants and Corresponding AQI

Cities	Days	PM_{10}	$PM_{2.5}$	NO_2	SO_2	AQI_{mean}
Ghaziabad	Week 1	92.169	96.797	56.169	23.347	88.223
	Week 2	84.009	73.300	43.069	19.132	70.982
	Week 3	84.007	73.307	43.069	19.216	71.010
	Week 4	84.008	73.307	43.069	19.244	71.019
Delhi	Week 1	264.509	135.737	23.425	23.633	132.997
	Week 2	240.739	97.698	21.594	20.494	109.996
	Week 3	240.323	97.797	21.204	20.446	109.87
	Week 4	240.069	97.797	21.145	20.446	109.799
Kolkata	Week 1	38.713	20.863	5.0413	3.0169	20.889
	Week 2	37.588	18.435	4.5062	2.907	19.395
	Week 3	37.587	18.435	4.504	2.907	19.394

	Week 4	37.587	18.435	4.505	2.907	19.394
Hyderabad	Week 1	121.433	25.460	13.777	14.635	49.845
	Week 2	112.192	29.503	15.167	11.841	48.781
	Week 3	112.188	29.516	15.107	11.401	48.629
	Week 4	112.188	29.516	15.130	12.645	49.025
Cochin	Week 1	02.129	03.177	19.062	12.496	11.718
	Week 2	03.027	03.557	18.393	13.018	12.054
	Week 3	02.927	03.495	18.393	12.579	11.866
	Week 4	02.897	03.361	18.393	11.877	11.584

From the table 59, the difference between the observed values and the forecasted values is seen. The main reason for the high forecast error is the unlocking process in India. This is different in between cities according to the condition of COVID-19 pandemic there. The use of Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) of the forecast errors for the comparison of the values. Since the values according to the conditions of the lockdown has been predicted, in some cases, the forecasted values are greater than the observed values because the values of the pollutants only depend on the unlocking process by each state governments.

Table 59: Forecast Errors for Predicted Values of Pollutants

Cities	Pollutants	Weeks	Observed vale	Forecasted vale	Forecast error	MAE	MAPE
Ghaziabad	PM_{10}	Week 1	45.5	92.169	46.669	40.470	46.3687
		Week 2	79.19	84.009	4.819		
		Week 3	158.97	84.007	-74.963		
		Week 4	119.44	84.008	-35.432		
	$PM_{2.5}$	Week 1	15.36	96.797	-81.437	50.445	238.619
		Week 2	26.74	73.3	-46.56		
		Week 3	48.25	73.307	-25.057		
		Week 4	24.58	73.307	-48.727		
	NO_2	Week 1	11.33	56.169	-44.839	29.706	204.054
		Week 2	14.57	43.069	-28.499		
		Week 3	21.34	43.069	-21.729		
		Week 4	19.31	43.069	-23.759		
	SO_2	Week 1	9.15	23.347	14.197	9.1647	90.0558
		Week 2	9.25	19.132	9.882		
		Week 3	13.69	19.216	5.526		
		Week 4	12.19	19.244	7.054		
Delhi	PM_{10}	Week 1	57.69	264.509	-206.81	141.63	171.625
		Week 2	104.62	240.739	-136.11		
		Week 3	159.79	240.323	-80.533		
		Week 4	97	240.069	-143.06		
	$PM_{2.5}$	Week 1	29.31	135.737	-106.427	61.65	177.33
		Week 2	47.06	97.698	-50.638		
		Week 3	74.17	97.797	-23.627		
		Week 4	31.88	97.797	-65.917		
	NO_2	Week 1	10.58	23.425	-12.845	12.162	148.127
		Week 2	10.39	21.594	-11.204		
		Week 3	12.35	21.204	-8.854		
		Week 4	5.4	21.145	-15.745		
	SO_2	Week 1	6.61	23.633	-17.023	11.904	143.049
		Week 2	12.46	20.494	-8.034		
		Week 3	8.3	20.446	-12.146		
		Week 4	10.03	20.446	-10.416		
Kolkata	PM_{10}	Week 1	38.46	38.713	0.253	7.2322	20.5407
		Week 2	49.88	37.588	12.292		
		Week 3	27.45	37.587	10.137		
		Week 4	31.34	37.587	6.247		
	$PM_{2.5}$	Week 1	18.07	20.863	2.793	6.9795	86.3421
		Week 2	15.19	18.435	3.245		

		Week 3	6.14	18.435	12.295		
		Week 4	8.85	18.435	9.585		
	NO_2	Week 1	3.06	5.0413	1.9813	1.2671	30.234
		Week 2	3.7	4.5062	0.8062		
		Week 3	6.72	4.504	2.216		
		Week 4	4.57	4.505	0.065		
	SO_2	Week 1	3.25	3.0169	0.2331	1.4225	60.5754
		Week 2	2	2.907	0.907		
		Week 3	1.21	2.907	1.697		
		Week 4	5.76	2.907	2.853		
Hyderabad	PM_{10}	Week 1	65.52	121.433	55.913	39.225	109.016
		Week 2	106.73	112.192	5.462		
		Week 3	25.68	112.188	86.508		
		Week 4	103.17	112.188	9.018		
	$PM_{2.5}$	Week 1	20.43	25.46	5.03	7.4742	66.6392
		Week 2	32.36	29.503	-2.857		
		Week 3	8.99	29.516	20.526		
		Week 4	31	29.516	-1.484		
	NO_2	Week 1	11.16	13.777	-2.617	2.0442	15.4536
		Week 2	17.86	15.167	2.693		
		Week 3	12.28	15.107	-2.827		
		Week 4	15.09	15.13	-0.04		
	SO_2	Week 1	7.05	14.635	-7.585	10.1	75.7253
		Week 2	30.11	11.841	18.269		
		Week 3	5.91	11.401	-5.491		
		Week 4	21.7	12.645	9.055		
Cochin	PM_{10}	Week 1	45.12	2.129	42.991	32.695	89.9912
		Week 2	14.32	3.027	11.293		
		Week 3	39.41	2.927	36.483		
		Week 4	42.91	2.897	40.013		
	$PM_{2.5}$	Week 1	15.74	3.177	12.563	14.395	78.1826
		Week 2	9.57	3.557	6.013		
		Week 3	22.56	3.495	19.065		
		Week 4	23.3	3.361	19.939		
	NO_2	Week 1	22.34	19.062	3.278	10.545	327.816
		Week 2	27.53	18.393	9.137		
		Week 3	5.38	18.393	-13.013		
		Week 4	1.64	18.393	-16.753		
	SO_2	Week 1	12.02	12.496	-0.476	6.213	30.5329
		Week 2	20.3	13.018	7.282		
		Week 3	20.92	12.579	8.341		
		Week 4	20.63	11.877	8.753		

3.3.2. Optimal allocation of limited resources

3.3.2.1 Saturation of the health-care systems

Kerala

The statistical tool "Time series" is used to fit the model. The time series plot for Kerala is given in Figure 129. The ARIMA model used is the one with least AIC (Akaike's information criteria).

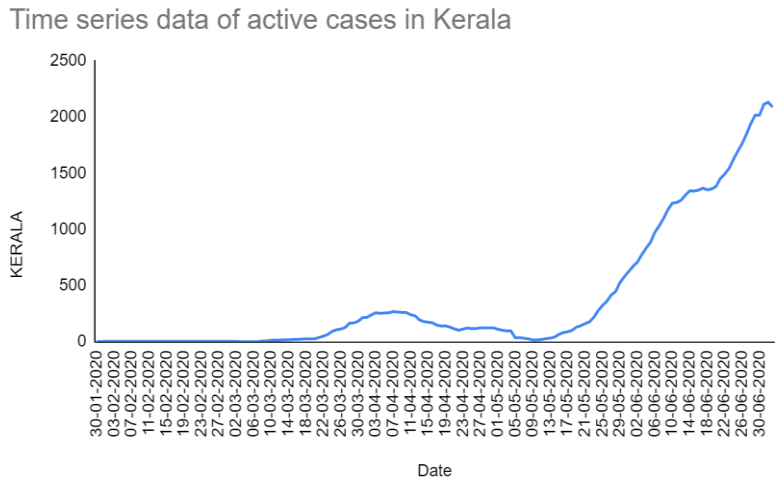


Figure 129: Time series data of active Cases of Kerala

The growth of active cases in Kerala from 30th January 2020 to 1st September 2020 is plotted in figure. Initially the no. of active cases gradually increases, then decreases and approaches the x-axis. Later the slope of the curve rises and reaches a peak of 2122 cases by July 5th, then decreases and arrives at a number of 1968 cases during the end of the study.

The Autocorrelation and partial autocorrelation graphs of undifferenced data of Kerala are given below (Figure 130).

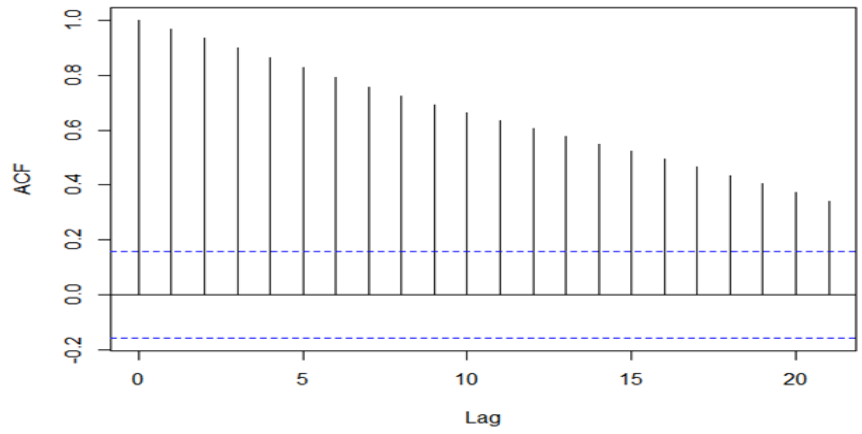


Figure 130: ACF plot of Kerala

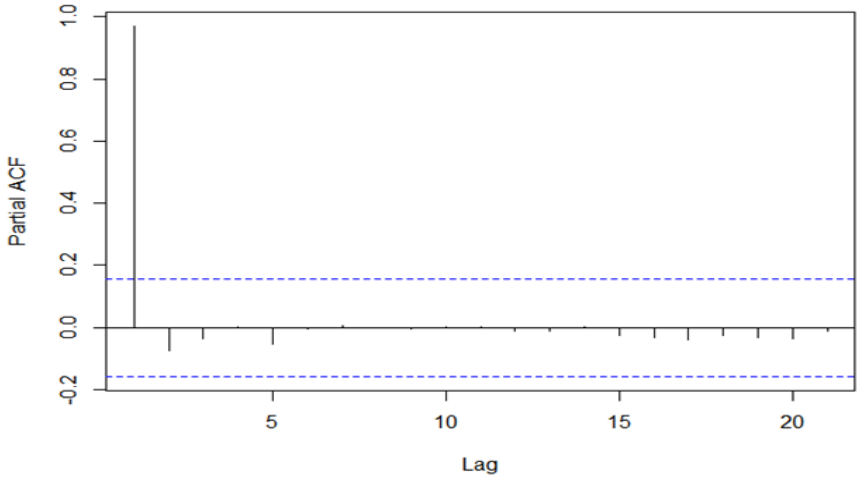


Figure 131: PACF plot of Kerala

Considering the plots of ACF and PACF, and also using the function auto. Arima it was observed that the model of Kerala is (2,2,2) (Figure 131). Figure 132 represents the Autocorrelation function of residuals of Kerala. The model with least AIC value is selected.

For model (2,2,2)

$Z_t=(\theta_1 a_t-a_t-\theta_1-+\theta_2 a_t-2\theta_2-4+3\Phi_2)/\Phi_2$ (6)

Φ_1 =-0.2300, Φ_2=-0.5816, θ_1=-0.3869, θ_2 = 0.4148

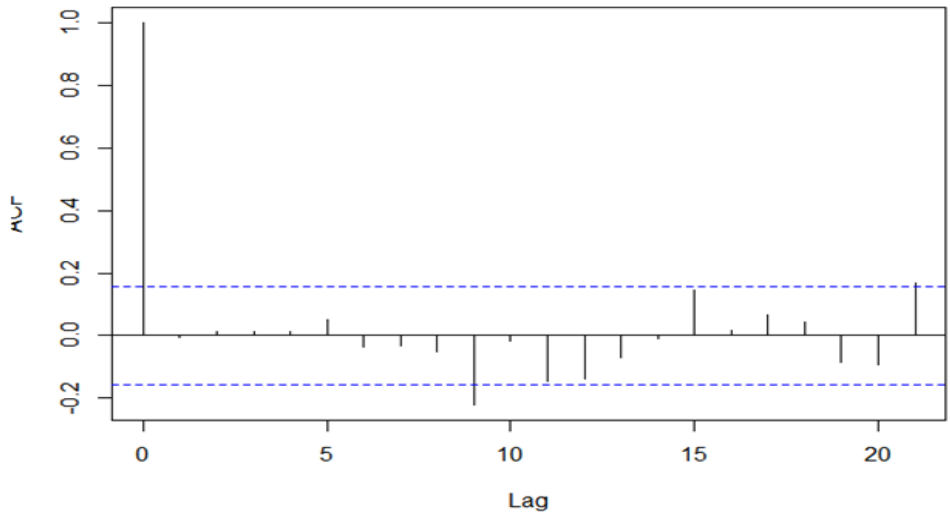


Figure 132: Autocorrelation function of residuals of Kerala

Table 60: Box-Ljung test of Kerala

X^2	0.0073862
Df	1
Significance level	0.9315

From the ACF and PACF graphs, it was seen that the residuals are uncorrelated. Box-Ljung test was used to confirm it. The p-value is above the 0.05 hence hypothesis of uncorrelation is accepted (Table 60).

Table 61: Shapiro-Wilk normality test of Kerala

W	0.8483
Significance level	2.083e-11

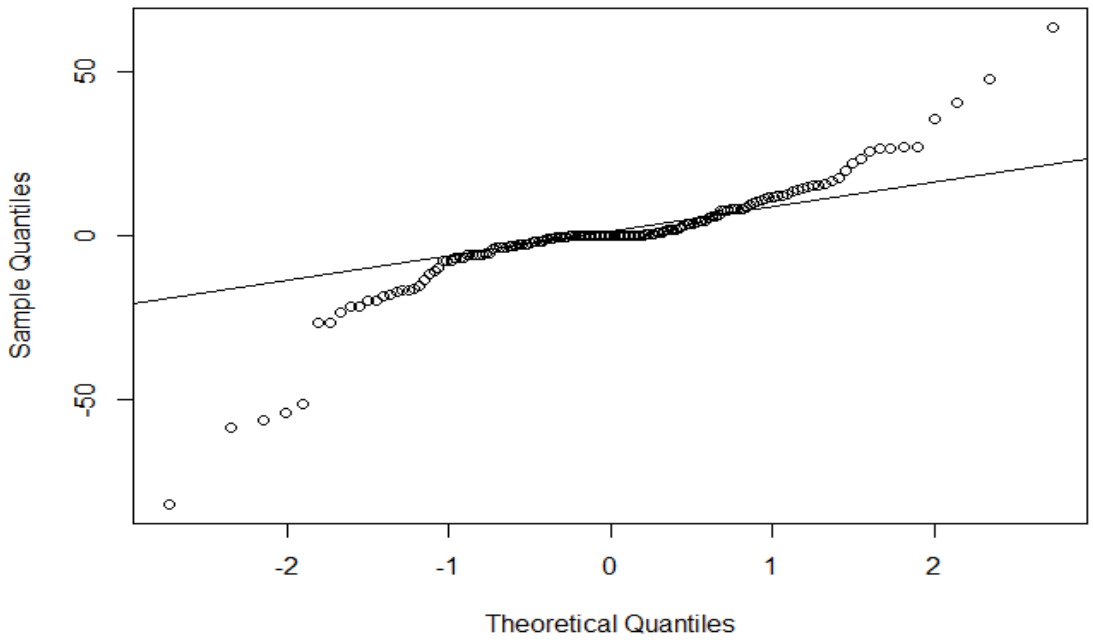


Figure 133: QQ Plot of residuals

Since the points cluster around a line, it is normal. The normality of errors is confirmed using the QQ-plot of residuals (Figure 133) and the Shapiro-Wilk test. Referring to the Shapiro-Wilk table (Table 61), it was found that the value of W to be substantially larger than the tabulated 25% point which is 0.8483. Thus, there is no evidence for the test for non-normality of residuals.

In-Sample forecasting

To check the validity of the proposed model in forecasting consider the data and forecast from 4th July to 10th July (Table 62).

Table 62: Forecasting active cases of Kerala using the (2,2,2) model from 4th July to 10th July

Day	The estimate of active cases	The lower limit of the interval	The upper limit of the interval	Actual active case
04/07/2020	2107	2073	2140	2098
05/07/2020	2122	2065	2180	2131
06/07/2020	2103	2021	2185	2230
07/07/2020	2094	1978	2210	2254
08/07/2020	2103	1946	2259	2415
09/07/2020	2102	1852	2333	2609

Karnataka

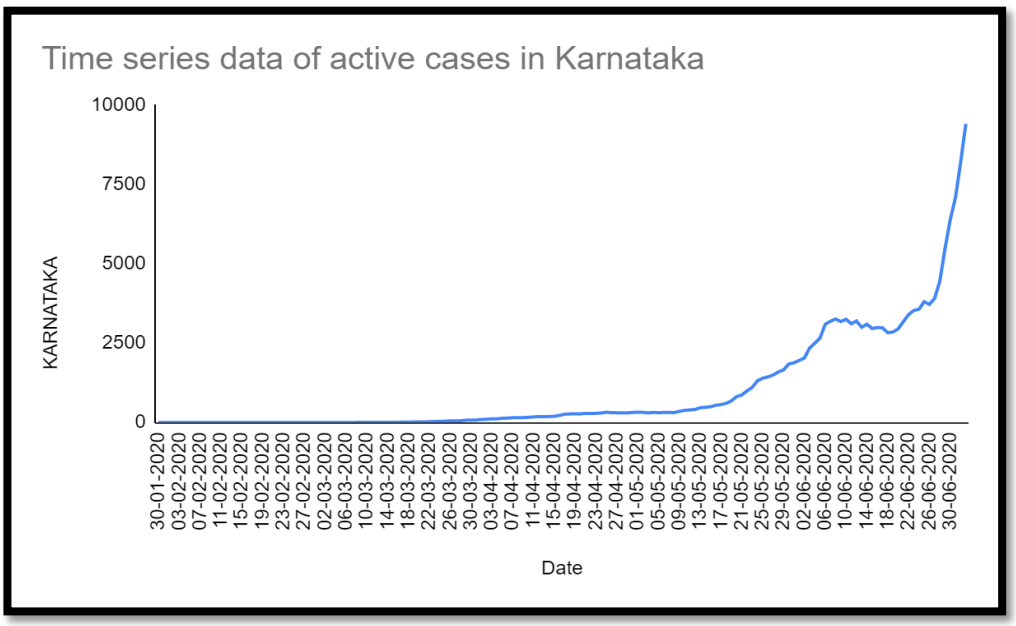


Figure 134: Active cases of Karnataka during the study period

Figure 134 shows active cases of Karnataka. The slope of the curve indicates that the active cases increase throughout the study. There were 80833 active cases in Karnataka on 1st September 2020. The shape of the curve indicates the gradual increase in the numbers in the upcoming months

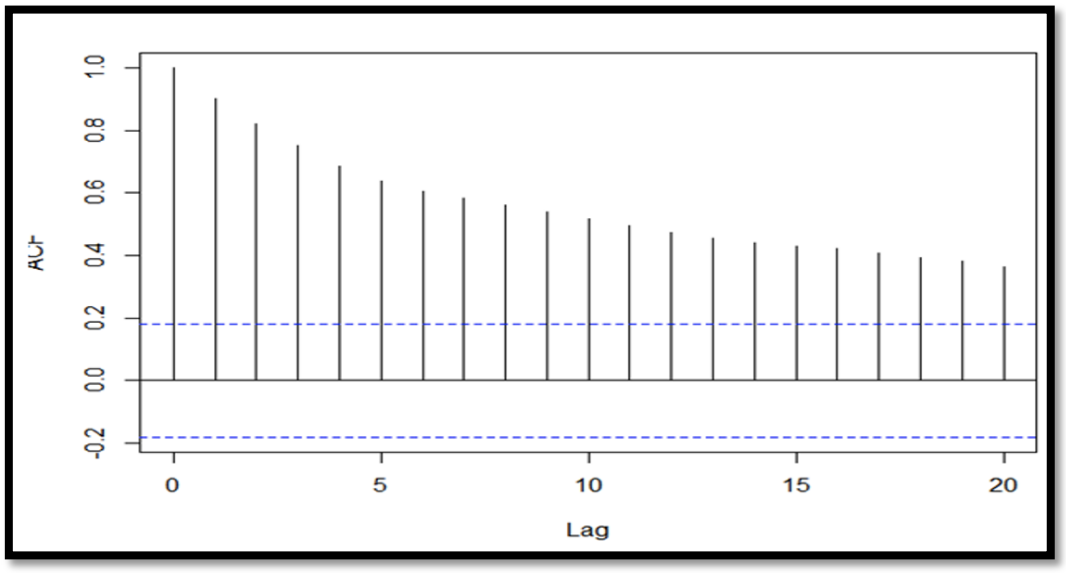


Figure 135: Autocorrelation function of Karnataka

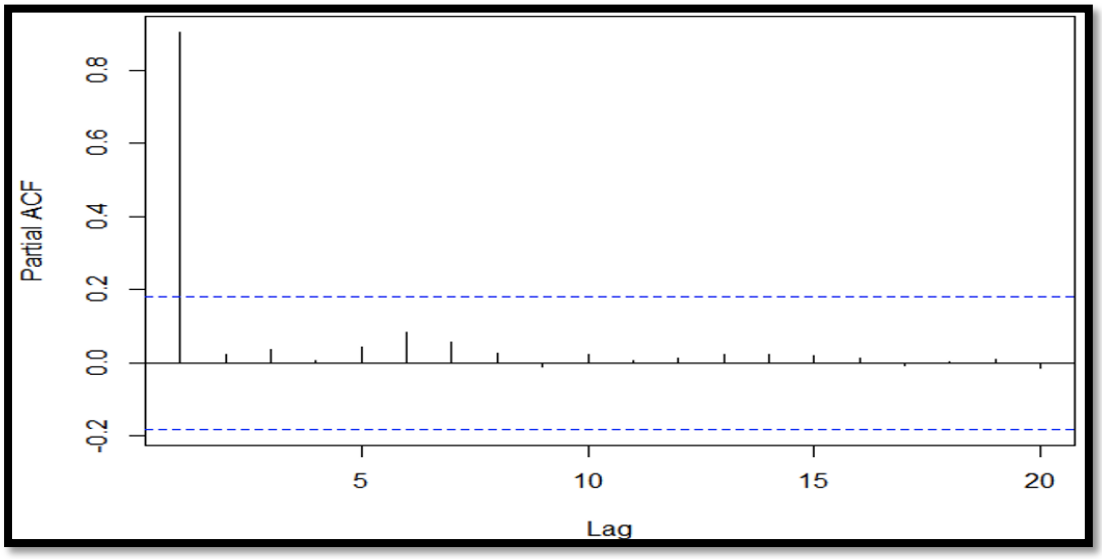


Figure 136: Partial autocorrelation function of Karnataka

Considering the plots of ACF (Figure 135) and PACF (Figure 136), and also using the function auto. ARIMA, it was observed that the model of Karnataka is (1,2,0). Figure 137 represents the Autocorrelation function of residuals of Karnataka. The model with least AIC value is selected.

For model (1,2,0)

$Z_t=(2B-B^2-1)$ (7)

Φ_1 =-.03054

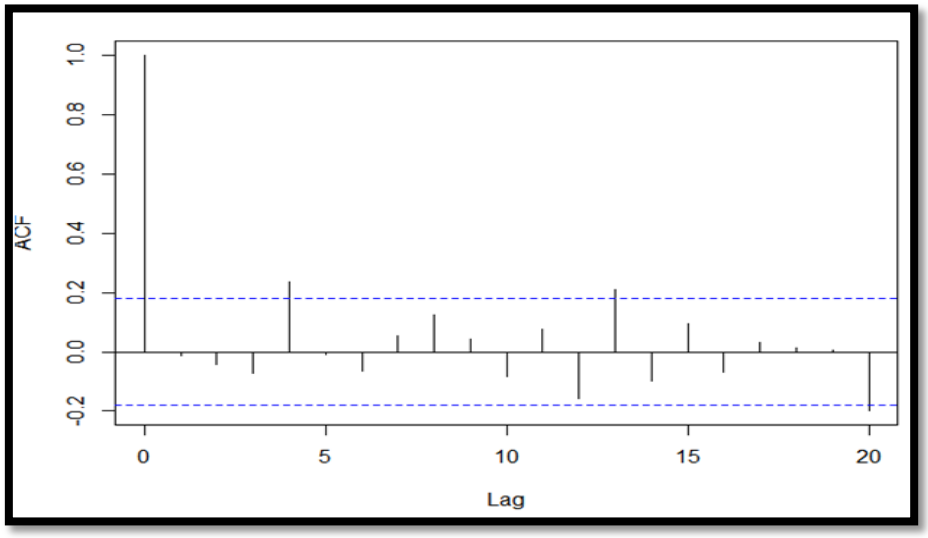


Figure 137: Autocorrelation function of residuals of Karnataka.

Table 63: Box-Ljung test of Karnataka

X^2	0.01935
Df	1
Significance level	0.8894

From the ACF graph (Figure 137) it was seen that the residuals are uncorrelated. Box-Ljung (Table 63) test can be used to confirm it. The p-value is above the 0.05 hence hypothesis of uncorrelation is accepted.

Table 64: Shapiro-Wilk normality test of Karnataka

W	0.78317
Significance level	7.422e-12

Referring to the Shapiro-Wilk table (Table 64), it was found that the value of W to be substantially larger than the tabulated 25% point which is 0.783. Thus, there is no evidence for the test for non-normality of residuals.

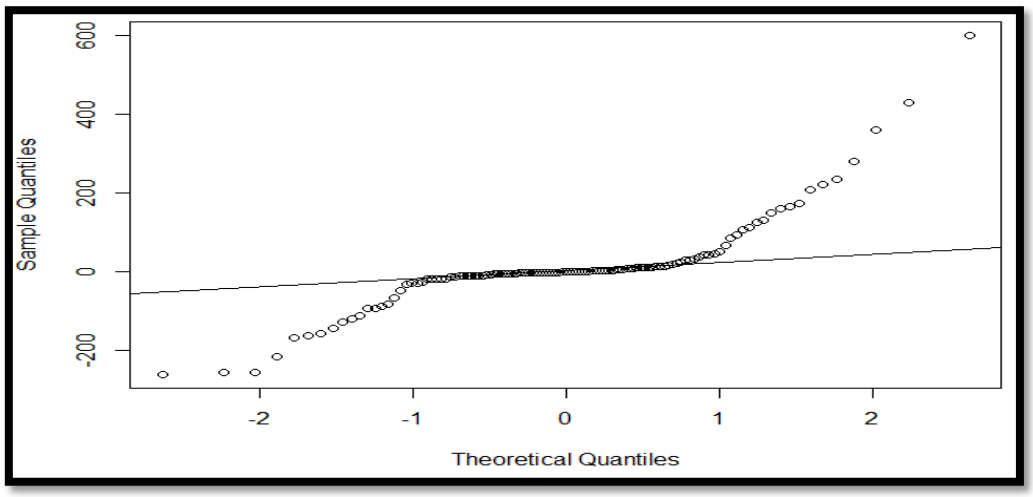


Figure 138: QQ Plot of residuals

QQ-plot of residuals is used to identify the normality of errors. Since the points cluster around a line, it is normal (Figure 138).

In-Sample forecasting

To check the validity of the proposed model in forecasting consider the data and forecast for the next few days (Table 65).

Table 65: Forecasting active cases of Karnataka using the (1,2,0) from 4th July to 10th July

Day	The estimate of active cases	The lower limit of the interval	The upper limit of the interval	Actual active cases
04/07/2020	10593	10366	10821	10612
05/07/2020	11786	11338	12234	11970
06/07/2020	12976	12255	13697	13255
07/07/2020	14166	13135	15198	14389
08/07/2020	15357	13979	16735	15301
09/07/2020	16547	14792	18302	16531

Tamil Nadu

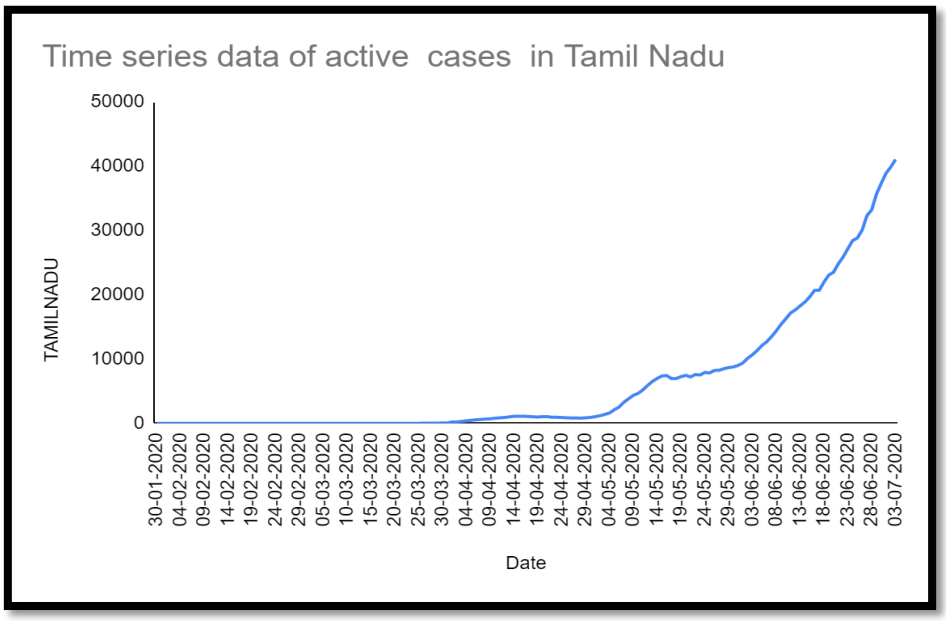


Figure 139: Active cases of Tamil Nadu during the study period

Figure conveys that the number of active cases is high. For Tamil Nadu, the active cases have a higher growth compared to Karnataka (Figure 139). There will be 111289 active cases in Tamil Nadu on 1st September 2020. The shape of the curve indicates the gradual increase in the numbers in the upcoming months. The peak of the cases on 04th August indicates that the state has to face massive growth of patients. The slope of the curve also indicates tremendous growth in the future.

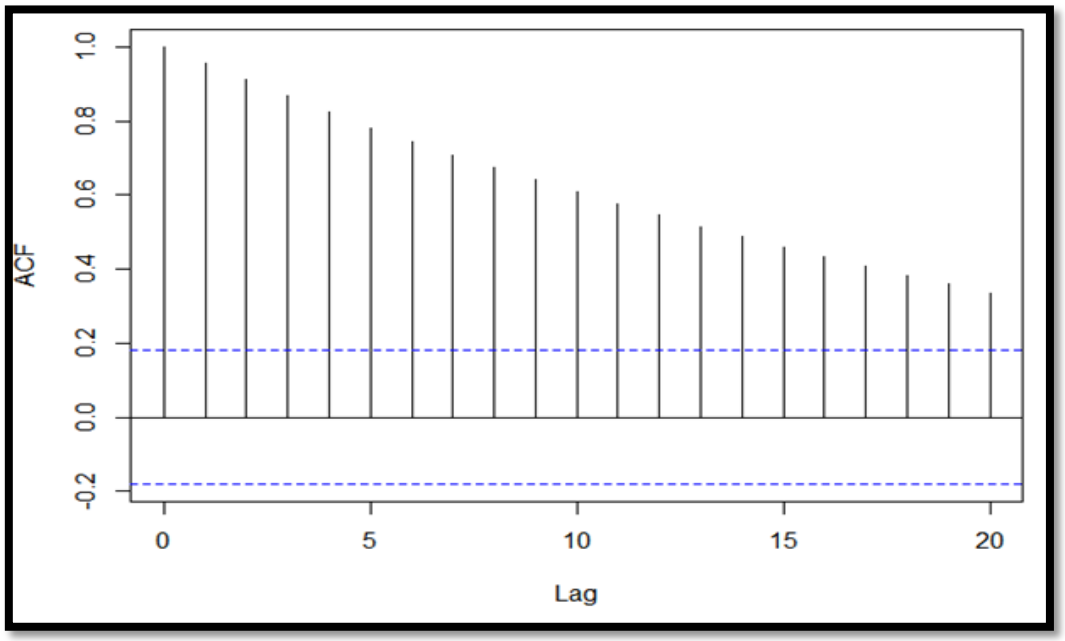


Figure 140: Autocorrelation function of Tamil Nadu

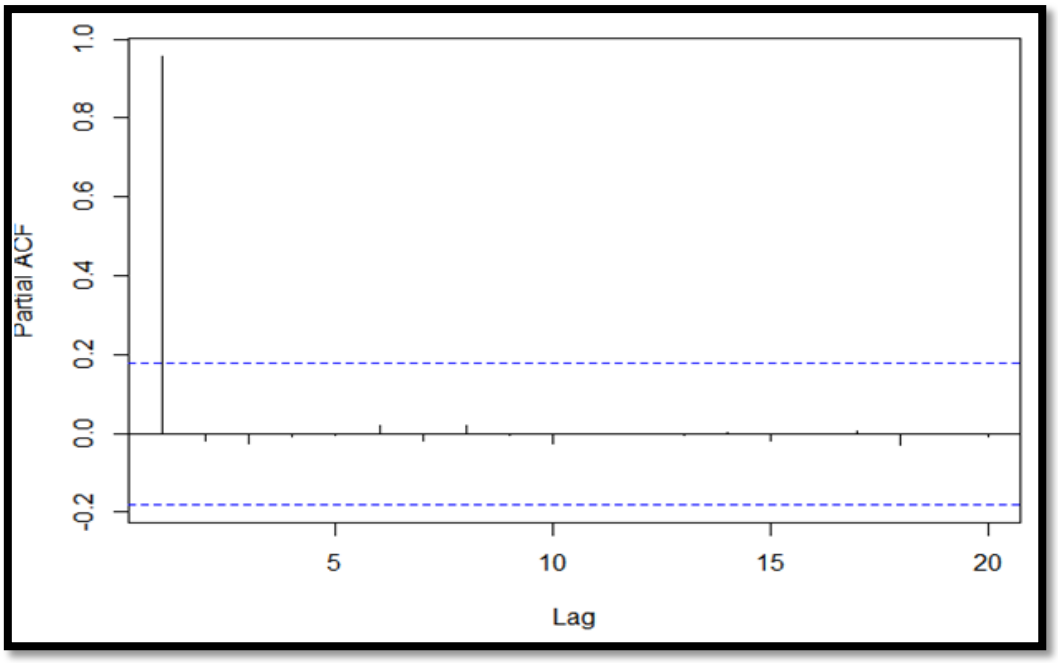


Figure 141: Partial Autocorrelation function of Tamil Nadu.

From the figures of ACF and PACF and also using the function auto (Figure 140, 141). Arima, find the model of Tamil Nadu is (0,2,1). Then selected the model with the least AIC value. For model (0,2,1), $a_t=\theta_1/(1-\theta_1)$ (5) θ_1= -0.9146, a_t=10.7where Z_t is the time series and a_t is the white noise

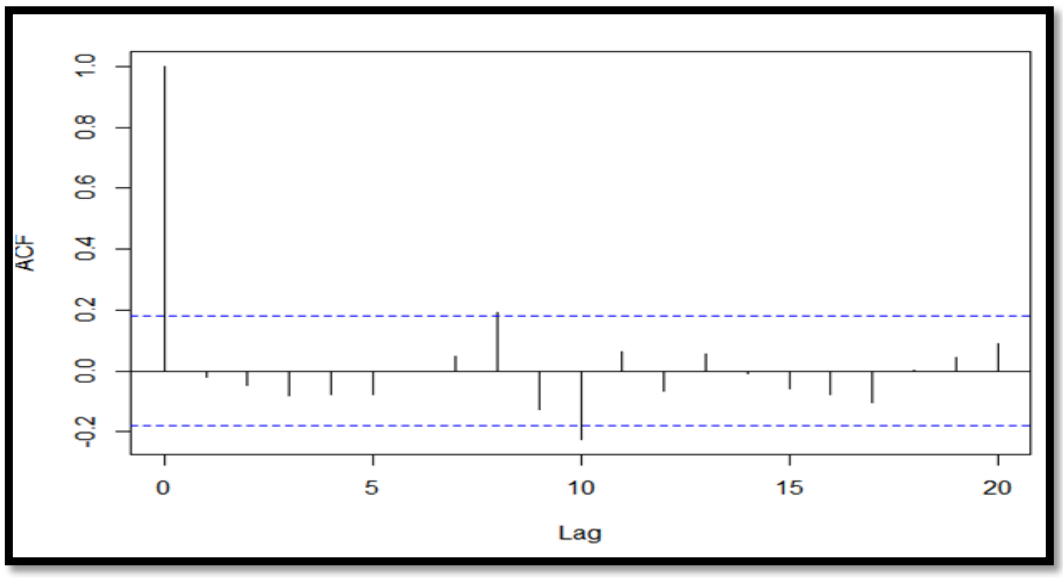


Figure 142: Autocorrelation functions of residuals of Tamil Nadu

Table 66: Box-Ljung test of Tamil Nadu

X^2	0.060709
Df	1
Significance level	0.8054

The residuals of the series are uncorrelated which was proved by test results (Figure 142). The output p-value of 0.8054 does not reject the null hypothesis of uncorrelated errors.

Table 67: Shapiro-Wilk normality test of Tamil Nadu

W	0.83731
Significance level	4.016e-10

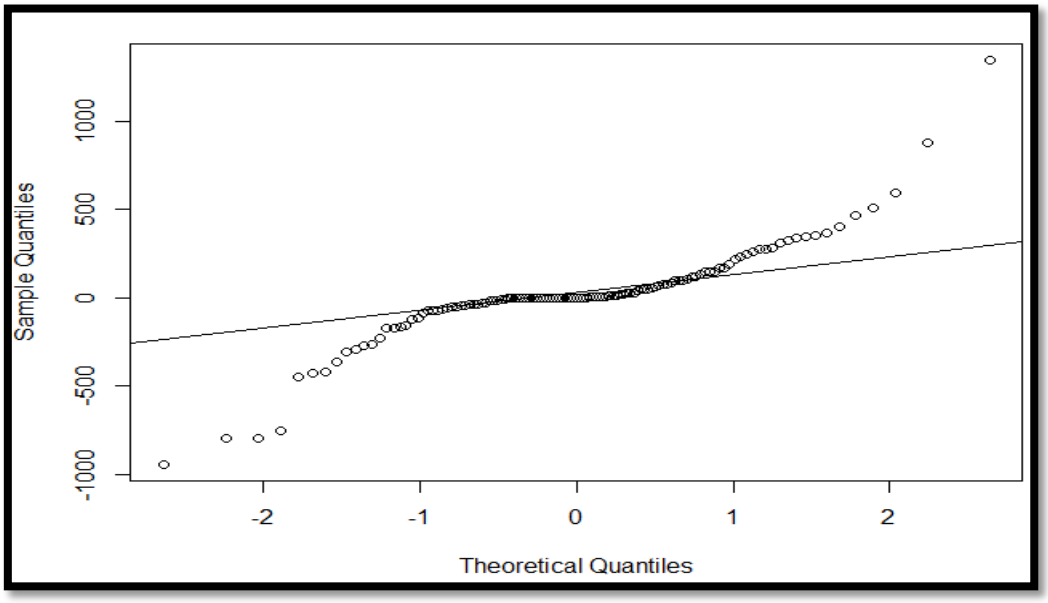


Figure 143: QQ Plot of residuals of Tamil Nadu

The normality of errors is confirmed using the QQ-plot of residuals and the Shapiro-Wilk test (Figure 143).

Referring to the Shapiro-Wilk table (Table 67), it was found that the value of W to be substantially larger than the tabulated 25% point which is 0. 83731.Thus there is no evidence for the test for non-normality of residuals.

In-Sample forecasting

To check the validity of the proposed model in forecasting consider the data and forecast for the next few days (Table 68).

Table 68: Forecasting active cases of Tamil Nadu using (0,2,1) model from 4th July to 10th July

Day	The estimate of active cases	The lower limit of the interval	The upper limit of the interval	Actual active cases
04/07/2020	42273	41726	42821	42958
05/07/2020	43396	42521	44272	44959
06/07/2020	44584	43298	45870	46863
07/07/2020	45757	43965	47549	46836
08/07/2020	46919	44597	49242	45842
09/07/2020	48095	45188	51001	46483

Maharashtra

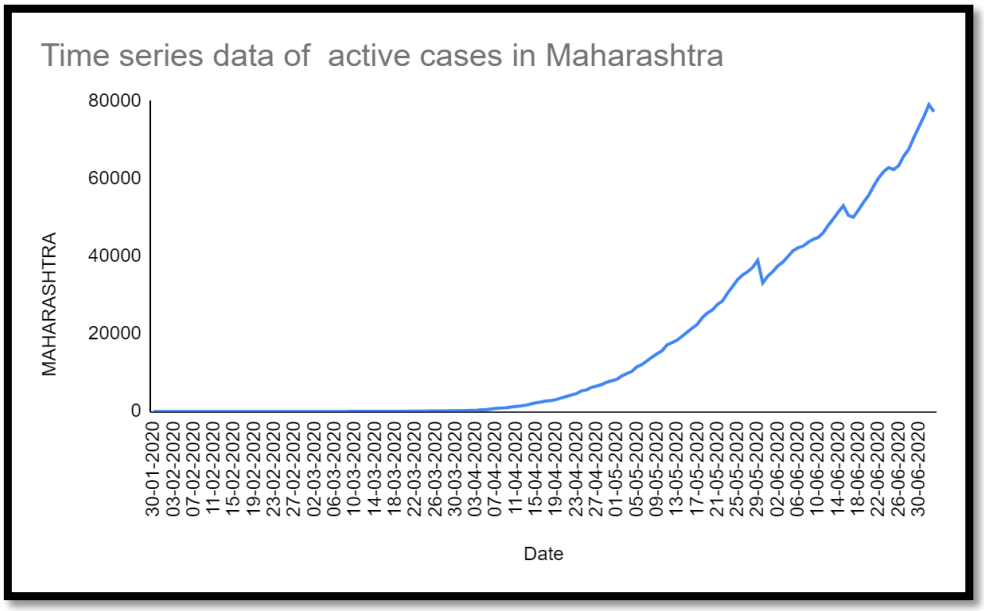


Figure 144: Active cases of Maharashtra

The higher slope of the line indicates a large infection rate (Figure 144). There will be 162505 active cases in Tamil Nadu on 1st September 2020. Furthermore, the situation of Maharashtra is not different but more complicated than other states. The shape of the curve indicates the gradual increase of the numbers in the upcoming months. The peak of the cases on 15th June 2020 is a warning that the state has to face massive growth of patients. Maharashtra is in troublesome condition. State and central government should take uncompromising measures to fight against coronavirus disease.

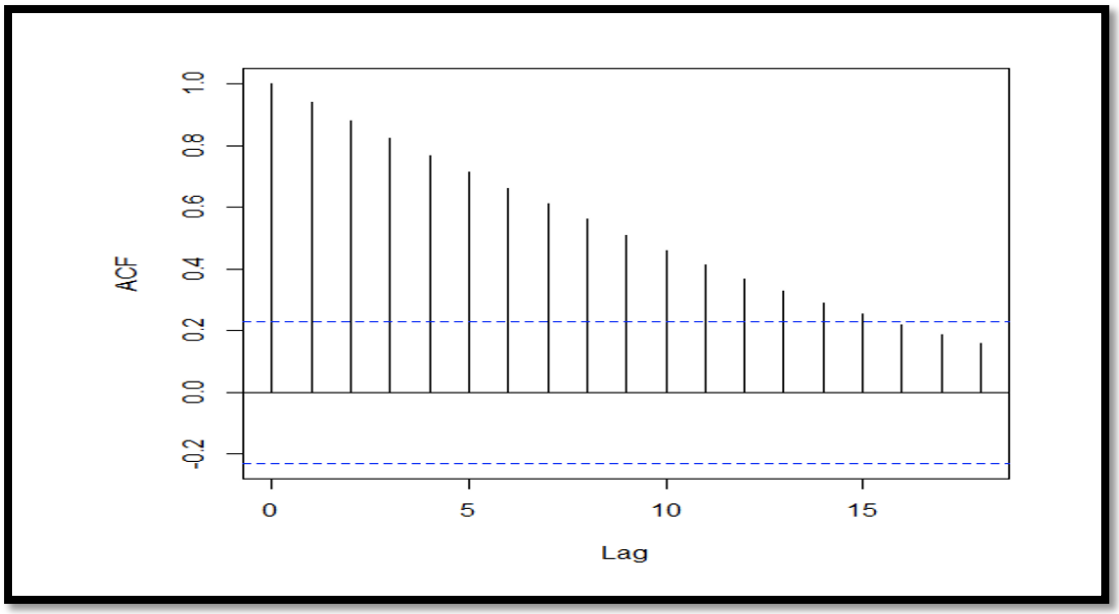


Figure 145: Autocorrelation Function of Maharashtra.

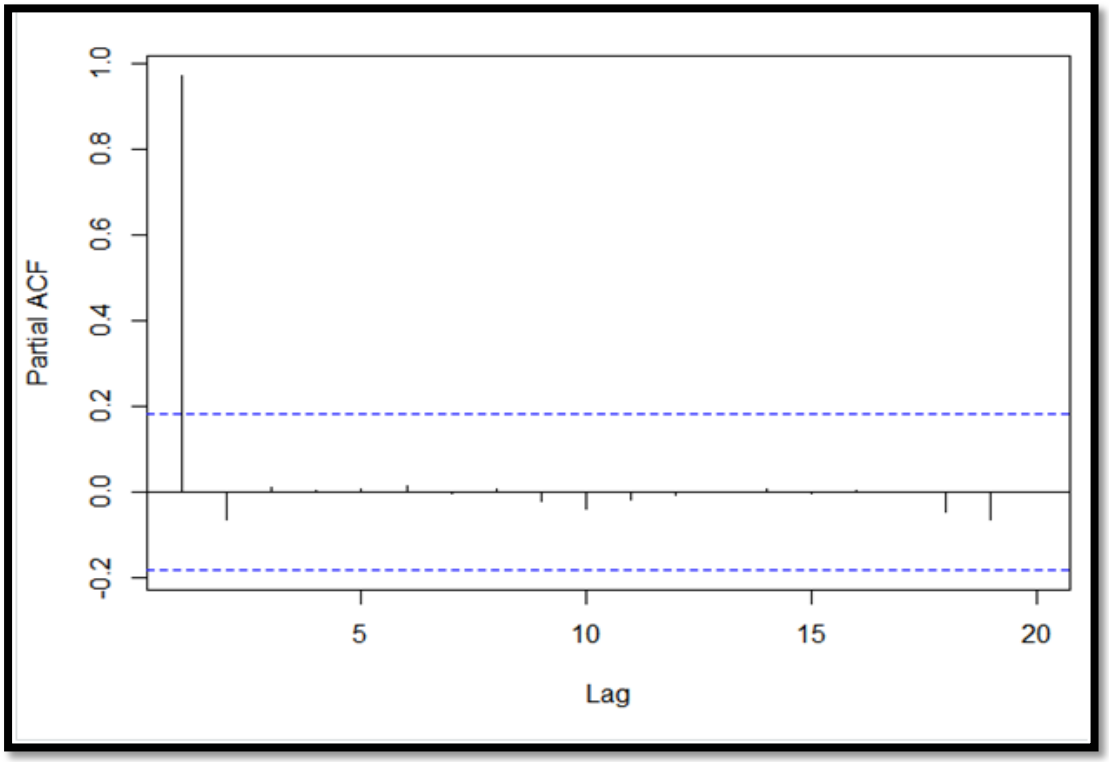


Figure 146: Partial Autocorrelation function of Maharashtra

From the figures of ACF and PACF and also using the function auto (Figure 145, 146). ARIMA find the model of Tamil Nadu is (2,2,0). Then selected the model with the least AIC value. Hence the identified model for the data is ARIMA (2,2,0) is given by

For model (2,2,0)

$a_t=4(\Phi_1-1)$

where $\Phi_1=-0.7515$

$a_t=4(-0.7515-1) =-1.7515$

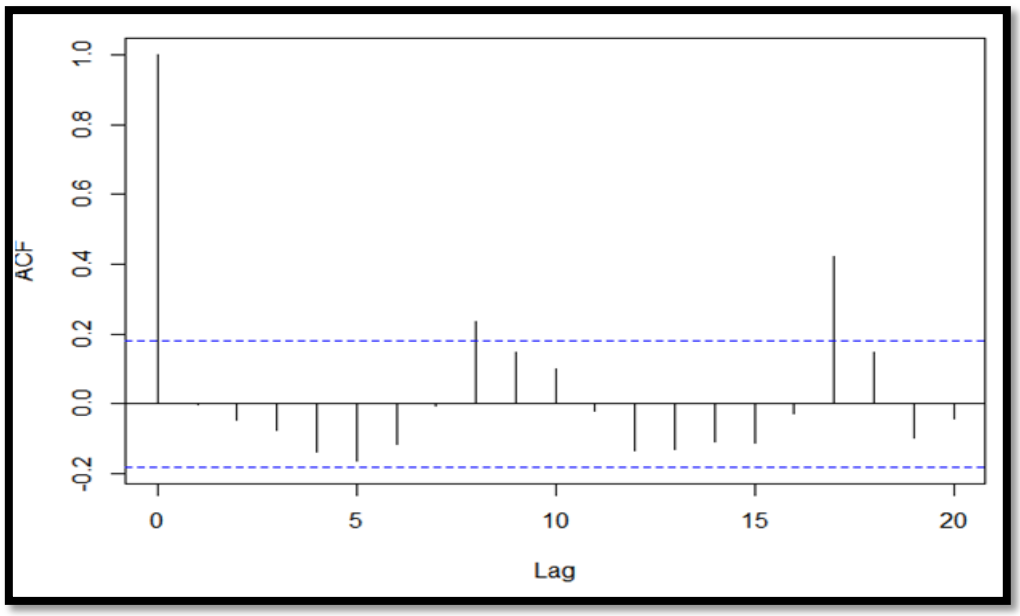


Figure 147: Autocorrelation function of residuals of Maharashtra

From the ACF (Figure 147) and it was seen that the residuals are uncorrelated. Box-Ljung test was used to confirm it (Table 69).

Table 69: Box-Ljung test of Maharashtra

X^2	0.00019552
Df	1
Significance level	0.9888

The output p value (0.9888) does not reject the null hypothesis of uncorrelated errors.

Table 70: Shapiro-Wilk normality test of Maharashtra

W	0.575932
Significance level	<2.2e4e-16

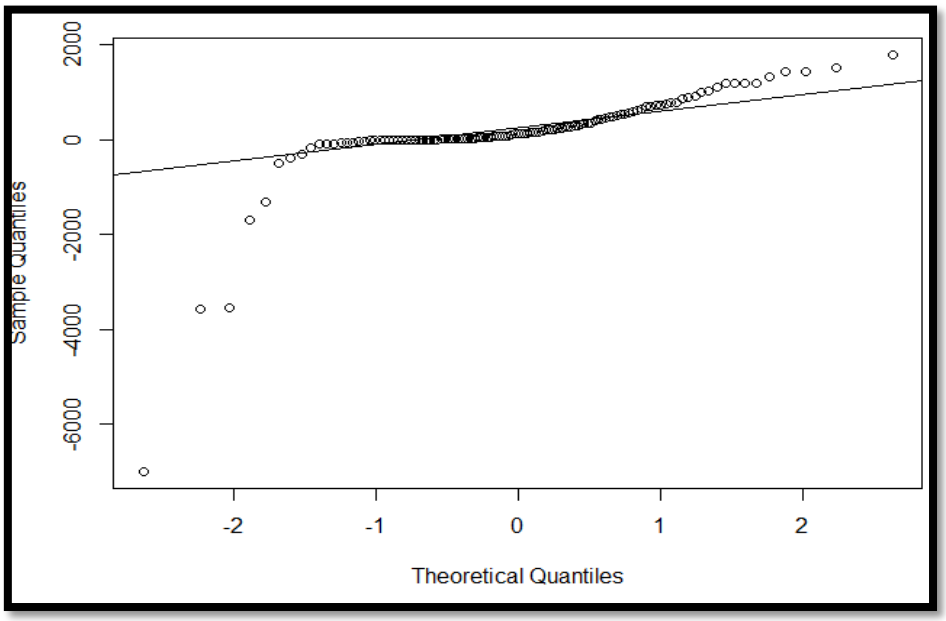


Figure 148: QQ Plot of residuals of Maharashtra

The normality of errors is confirmed using the QQ-plot of residuals (Figure 148).

Referring to the Shapiro-Wilk table (Table 70), it was found that the value of W to be substantially larger than the tabulated 25% point. Thus, there is no evidence for the test for non-normality of residuals.

In-Sample forecasting

To check the validity of the proposed model in forecasting consider the data and forecast from the 4th July to 10th July (Table 71).

Table 71: Forecasting active cases of Maharashtra using the (2,2,0) model from 4th July to 10th July

Day	An estimate of active cases	The lower limit of the interval	The upper limit of the interval	Actual active cases
04/07/2020	78696	76768	80624	79927
05/07/2020	80116	77272	82961	83311
06/07/2020	81537	77906	85168	86057
07/07/2020	82957	78593	87322	87699
08/07/2020	84378	79305	89451	89313
09/07/2020	85798	80027	91569	91084

Table 72: Weekly prediction of active cases for four states during the study period

Date	Kerala	Karnataka	Maharashtra	Tamil Nadu
23-07-2020	2078	27262	95742	56286
31-07-2020	2045	42738	117049	73840
08-08-2020	2026	52262	128413	83202
16-08-2020	2006	61785	139777	92565
24-08-2020	1987	71309	151141	101927

The weekly predictions of active COVID-19 cases presented in Table 72 offer valuable insights into the expected trends for four prominent states—Kerala, Karnataka, Maharashtra, and Tamil Nadu—during the study period from July 23, 2020, to August 24, 2020. This table provides a detailed breakdown of forecasted active cases on specific dates. As of the initial date, July 23, 2020, the predicted active cases were 2078 for Kerala, 27262 for Karnataka, 95742 for Maharashtra, and 56286 for Tamil Nadu. Subsequent weekly projections allow for a dynamic understanding of how these numbers are anticipated to change, aiding in the comprehensive analysis of the pandemic's progression in these states.

Demand and capacity of hospital infrastructure facilities

A graph is obtained by plotting the demand of ICU, ventilator and beds on each day for four states (Table 149). By observing the graph on demand for facilities along

with their carrying capacity, it is possible to derive the conclusions on the capability of states to deal with COVID-19.

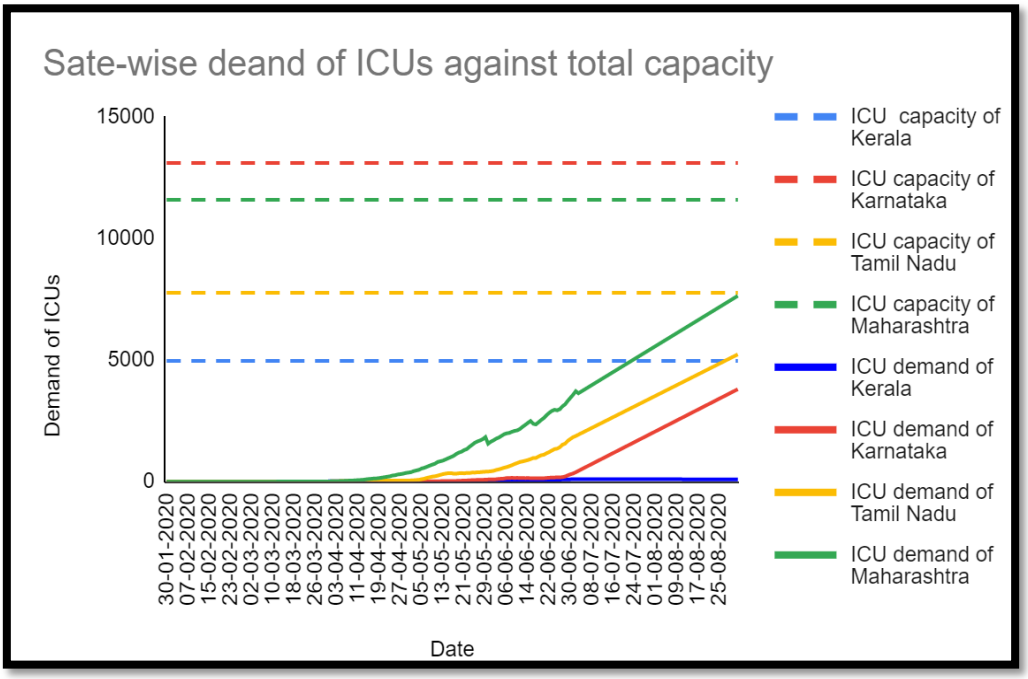


Figure 149: State-wise demand for ICU against total capacity

The ICU capacity of Kerala is closer to the x-axis since a maximum of 100 is needed. The demand for ICU by 1st September 2020 is 92 and the carrying capacity is 4791. For Karnataka the demand for ICU is 3799 and the carrying capacity is 13105 at the end of the study, and in Tamil Nadu, the demand for ICU will be 5231 by September 1st 2020 and the carrying capacity will be 7769. In Maharashtra on 1st September 2020 the demand for ICU beds will be 7637 and the carrying capacity is 11587.

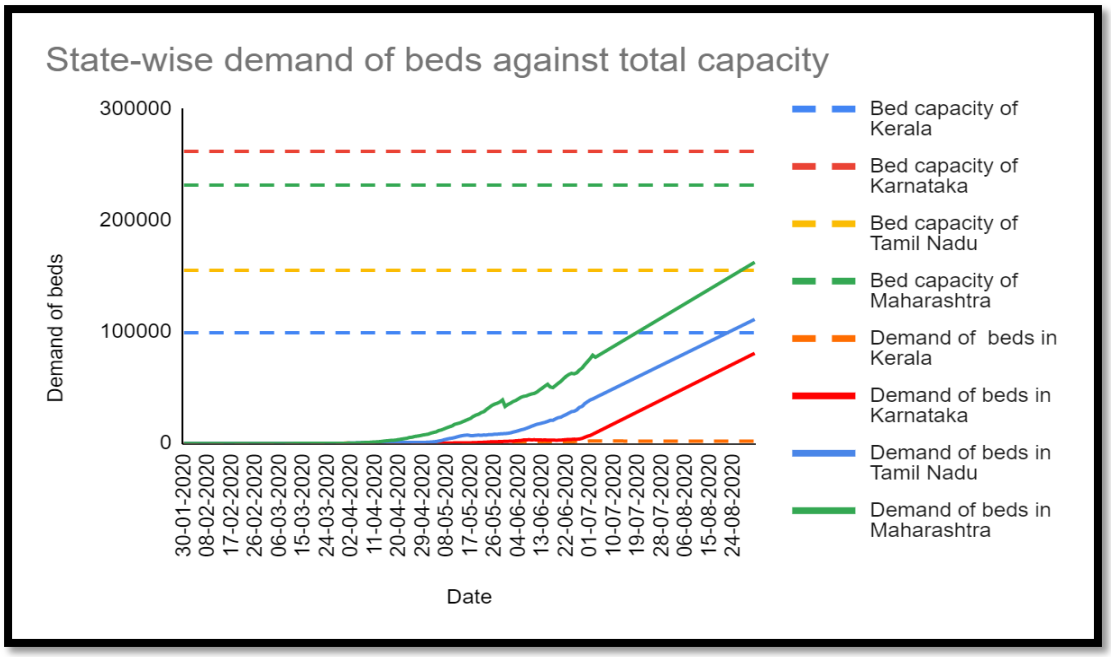


Figure 150: State-wise demand of beds against total capacity

In Kerala there is a demand of 1968 beds on September 1st and the carrying capacity is 99277. In Tamil Nadu, the demand for beds is 11289 and the carrying capacity is 155375. The demand for hospital beds in Karnataka is 80833 and the maximum capacity is 262109. At the end of the study the demand for beds in Maharashtra becomes 162505 and the carrying capacity is 231739 (Figure 150).

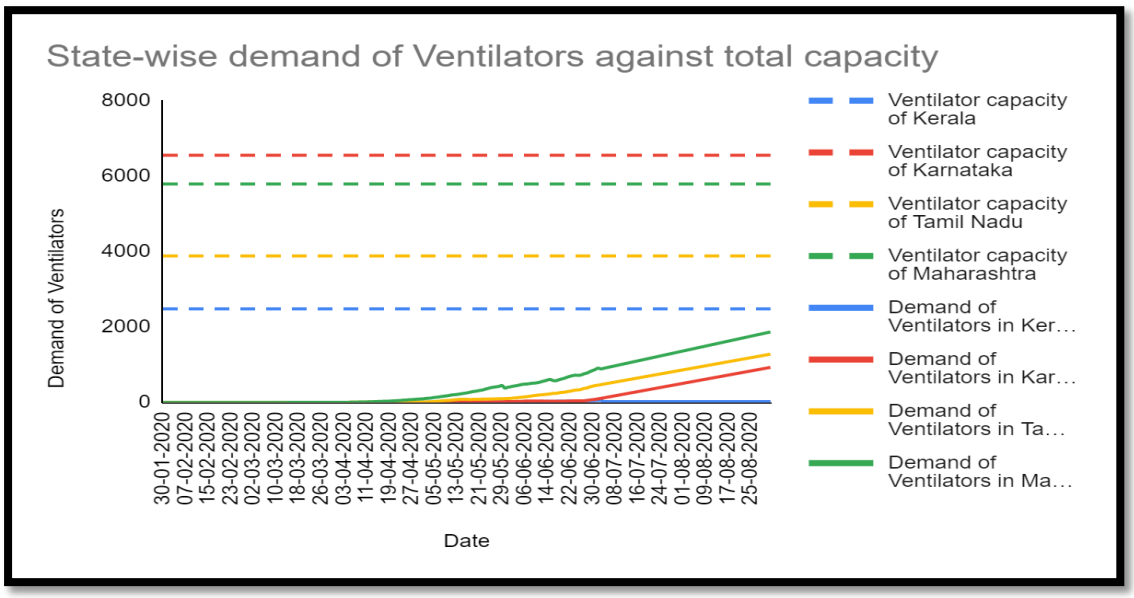


Figure 151: State Wise demand of ventilators against total ventilator capacity

In the case of ventilator also the demand will not exceed the carrying capacity. In Kerala the demand is 23 ventilators and the available capacity is 2481. In Tamil Nadu the demand is 1280 and the carrying capacity is 3884. In Karnataka the demand for ventilators is 930 and the carrying capacity is 6553. Also, in Maharashtra the demand is 1869 by the end of this study and the carrying capacity is 5793. Uncertainty in the duration of disease, the top utilization of facilities, etc raises challenges. The results are satisfactory, but in Tamil Nadu the demand for ICU, hospital beds and ventilators are a bit closer to the carrying capacities (Figure 151).

3.3.3. Impact Assessment Tool

3.3.3.1 Social impact

Among 800 respondents, 779 agreed to participate and completed the questionnaire. Out of the total respondents, 770 (98.8%) are aware about the precautionary measures against the pandemic. The precautionary measures include covering mouth while coughing and sneezing, avoid public gathering, washing hands and wearing mask. 86.1% of the people cover their mouth while coughing and sneezing, 77.02% usually avoid pubic gathering, 64.05% wash their hands frequently and 93.9% always wear mask while going out. Major sources of information include friends (18.6%), mass media (57%) and social media (24.4%).

The psychological impact, assessed by the Impact of Event Scale-Revised revealed a mean score of 23.55. Of all the respondents 429 (55.0%) reported a normal psychological impact, 135 (17.3%) reported a mild psychological impact, 54 (6.9%) reported a moderate psychological impact and 161 (20.6%) reported a severe psychological impact. The mental health assessed by the Depression Anxiety Stress

subscale revealed a mean score of 7.90, 5.73, and 9.09 respectively. For the Depression subscale 473 (60.7%) reported a normal mental health status, 100 (12.8%) reported a mild mental health status, 154 (19.7%) reported a moderate mental health status and 52 (6.6%) reported a severe mental health status. For the Anxiety subscale 537 (68.9%) reported a normal mental health status,52 (6.6%) reported a mild mental health status, 119 (15.2%) reported a moderate mental health status and 71 (9.1%) reported a severe mental health status. For the Stress subscale 473 (60.7%) reported a normal mental health status, 215 (27.5%) reported a mild mental health status, 69 (8.8%) reported a moderate mental health status and 22 (2.8%) reported a severe mental health status (Table 73, Figure 152).

Table 73: Summary of IES-R and Depression Anxiety Stress Scales

Range	Depression	Anxiety	Stress	IES-R
Normal	473 (60.71%)	537 (68.9%)	473 (60.7%)	429(55%)
Mild	100 (12.8%)	52 (6.6%)	215 (27.5%)	135 (17.3%)
Moderate	154 (19.7%)	119 (15.2%)	69 (8.8%)	54 (6.9%)
Severe	52 (6.6%)	71 (9.1%)	22 (2.8%)	161(20.6%)

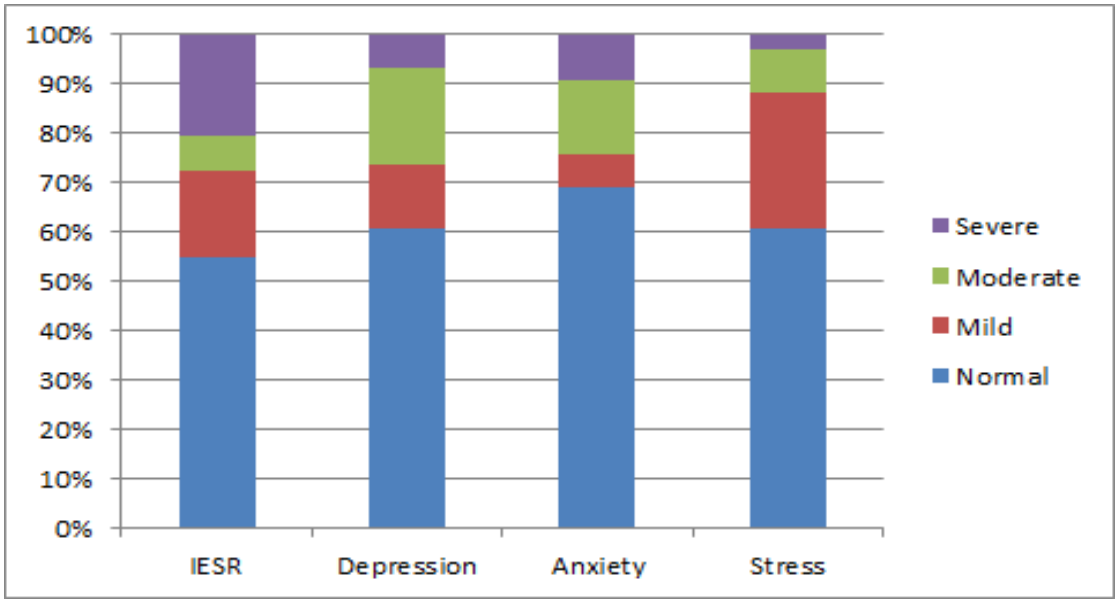


Figure 152: Summary of IES-R and Depression Anxiety Stress scale

Association between the IES-R scale and demographic variables during lockdown

The chi-square test is used to find an association between different demographic variables and the IES-R and Depression Anxiety & Stress Scale.

H_0: There is no association between various demographic variables and IES-R Scale.

H_1: There is an association between various demographic variables and IES-R Scale.

Table 74: Psychological Impact (IESR) by Socio-economic and Demographic Variables

Variable	Groups	IESR				Sig.
		Normal	Mild	Moderate	Severe	
Gender	Male	177	66	24	69	0.479
	Female	252	69	30	95	
	Total	429	135	54	161	
Age	18-25	317	104	46	132	0.161
	26-40	88	27	7	26	
	Above 40	24	4	1	3	
	Total	429	135	54	161	
Area of residence	Urban	132	53	13	66	0.023*
	Rural	297	82	41	95	
	Total	429	135	54	161	
Marital status	Single	351	117	47	141	0.245
	Married	78	18	7	20	
	Total	429	135	54	161	
Educational Qualification	Plus, two or below	29	15	3	11	0.594
	Graduate	211	63	31	82	
	Post Graduate	189	57	20	68	
	Total	429	135	54	161	
Source of information	Friends	87	17	6	35	0.042*
	Mass media	251	76	29	88	
	Social media	91	42	19	38	
	Total	429	135	54	161	
Social media exposure	Never	35	9	4	9	0.916
	Sometimes	91	34	16	37	
	Often	126	36	15	46	
	Very often	177	56	19	69	
	Total	429	135	54	161	
Covering mouth	Always	374	110	46	141	0.367
	Sometimes	55	25	8	20	
	Total	429	135	54	161	
Washing hands	Frequently	283	91	31	94	0.201
	Sometimes	146	44	23	67	
	Total	429	135	54	161	
Wearing mask	Yes	407	128	51	146	0.275
	No	22	7	3	15	
	Total	429	135	54	161	
Avoid public gathering	Always	347	95	38	120	0.031*
	Sometimes	82	40	16	41	
	Total	429	135	54	161	

Extra activities	Doing meditation or yoga	47	10	7	24	0.009*
	Using social media	69	31	16	28	
	Doing something I am good at	129	52	17	51	
	Spending time with loved ones	79	28	7	24	
	None	105	14	7	34	
	Total	429	135	54	161	
Chronic diseases	Yes	12	5	1	9	0.364
	No	417	130	53	152	
	Total	429	135	54	161	
Monthly family income	Below 5000	45	12	6	30	0.061
	5000-20000	126	50	21	57	
	20000-50000	146	49	13	46	
	50000-75000	52	12	6	11	
	75000 and above	60	12	8	17	
	Total	429	135	54	161	

* p value is less than 0.05.

Table 74 shows that the variables like area of residence, source of information, avoiding public gathering and extra activities are important factors for various levels of psychological impact among people during lockdown. Psychological impact of this significant variables associated with IES-R scale is presented in Figure 153.

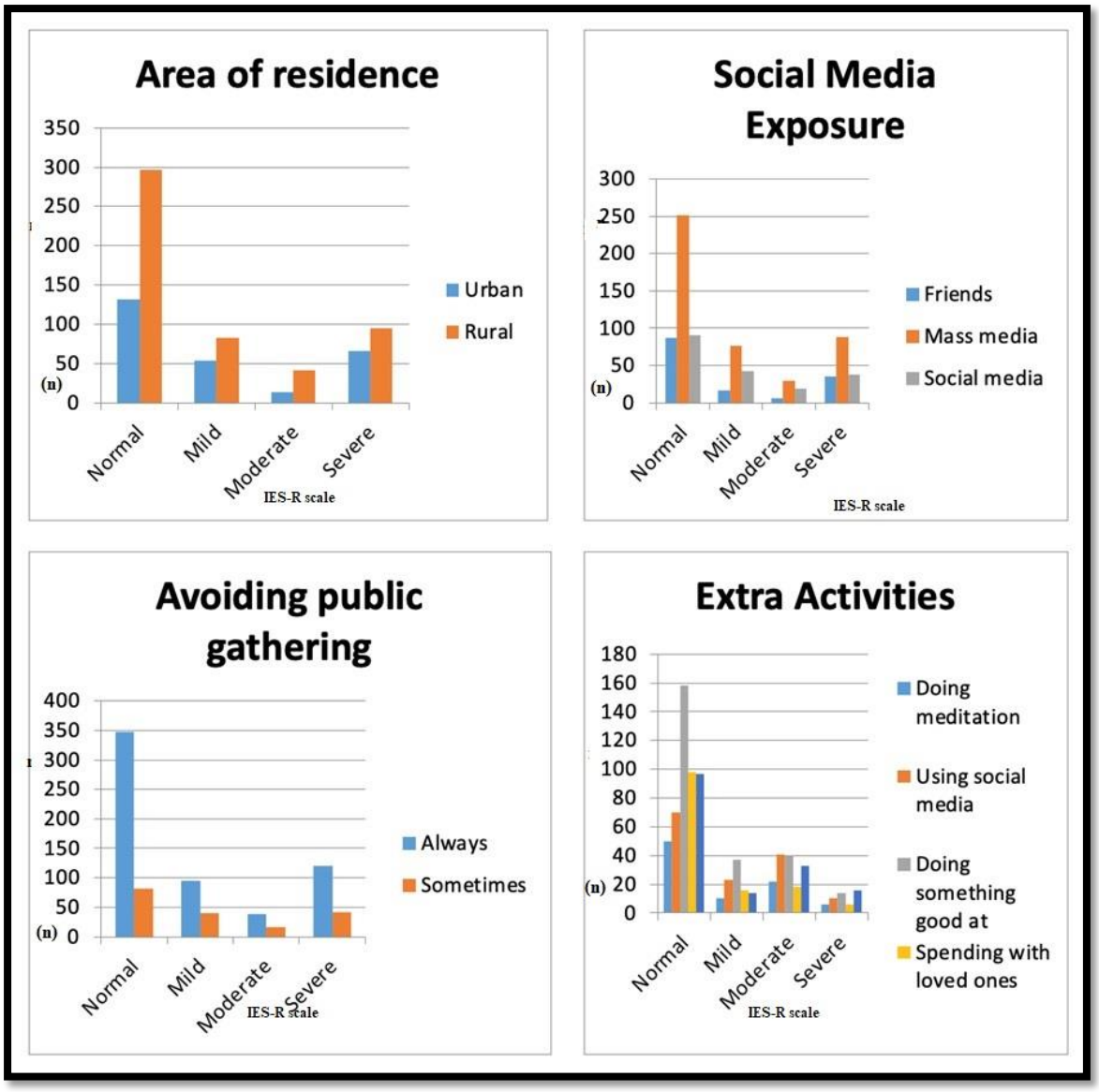


Figure 153: Psychological impact of significant variables associated with IES-R scale

Association between anxiety and demographic variables during lockdown

Table 75: Anxiety by Socio-economic and Demographic Variables

Variable	Groups	Anxiety				Sig.
		Normal	Mild	Moderate	Severe	
Gender	Male	226	19	59	32	0.349
	Female	311	33	60	39	
	Total	537	52	119	71	
Age**	18-25	400	41	101	57	0.348
	26-40	112	9	15	12	
	Above 40	25	2	3	2	
	Total	537	52	119	71	
Area of residence	Urban	176	20	46	22	0.521
	Rural	361	32	73	49	
	Total	537	52	119	71	
Marital status	Single	442	44	105	65	0.122
	Married	95	8	14	6	
	Total	537	52	119	71	
Educational Qualification**	Plus, two or below	38	2	12	6	0.467
	Graduate	268	31	51	37	
	Post Graduate	231	19	56	28	
	Total	537	52	119	71	
Awareness on precautionary measures	Yes	535	51	117	67	0.001*
	No	2	1	2	4	
	Total	537	52	119	71	
Source of information	Friends	101	7	20	17	0.349
	Mass media	311	27	72	34	
	Social media	125	18	27	20	
	Total	537	52	119	71	
Social media exposure	Never	39	5	6	7	0.045*
	Sometimes	112	15	31	20	
	Often	146	19	43	15	
	Very often	240	13	39	29	
	Total	537	52	119	71	
Covering mouth	Always	474	41	100	56	0.047*
	Sometimes	63	11	19	15	
	Total	537	52	119	71	
Washing hands	Frequently	348	37	67	47	0.211
	Sometimes	189	15	52	24	
	Total	537	52	119	71	

Wearing mask**	Yes	506	49	110	67	0.901
	No	31	3	9	4	
	Total	537	52	119	71	
Avoid public gathering	Always	419	39	89	53	0.804
	Sometimes	118	13	30	18	
	Total	537	52	119	71	
Extra activities	Doing meditation or yoga	59	7	11	11	0.056
	Using social media	84	14	32	14	
	Doing something I am good at	170	20	37	22	
	Spending time with loved ones	108	6	14	10	
	None	116	5	25	14	
	Total	537	52	119	71	
Chronic diseases**	Yes	13	2	4	8	0.002*
	No	524	50	115	63	
	Total	537	52	119	71	
Monthly family income	Below 5000	59	5	16	13	0.267
	5000-20000	167	14	51	22	
	20000-50000	184	21	31	18	
	50000-75000	60	5	8	8	
	75000 and above	67	7	13	10	
	Total	537	52	119	71	

Note: * p value is less than 0.05, ** Monte Carlo simulated value.

From table 75 it was seen that variables such as exposure to social media, covering mouth while coughing and sneezing, chronic diseases and awareness on precautionary measures are significant factors for anxiety among the study participants during this pandemic. The measure of this significant variables associated with anxiety during lockdown is presented in Figure 154.

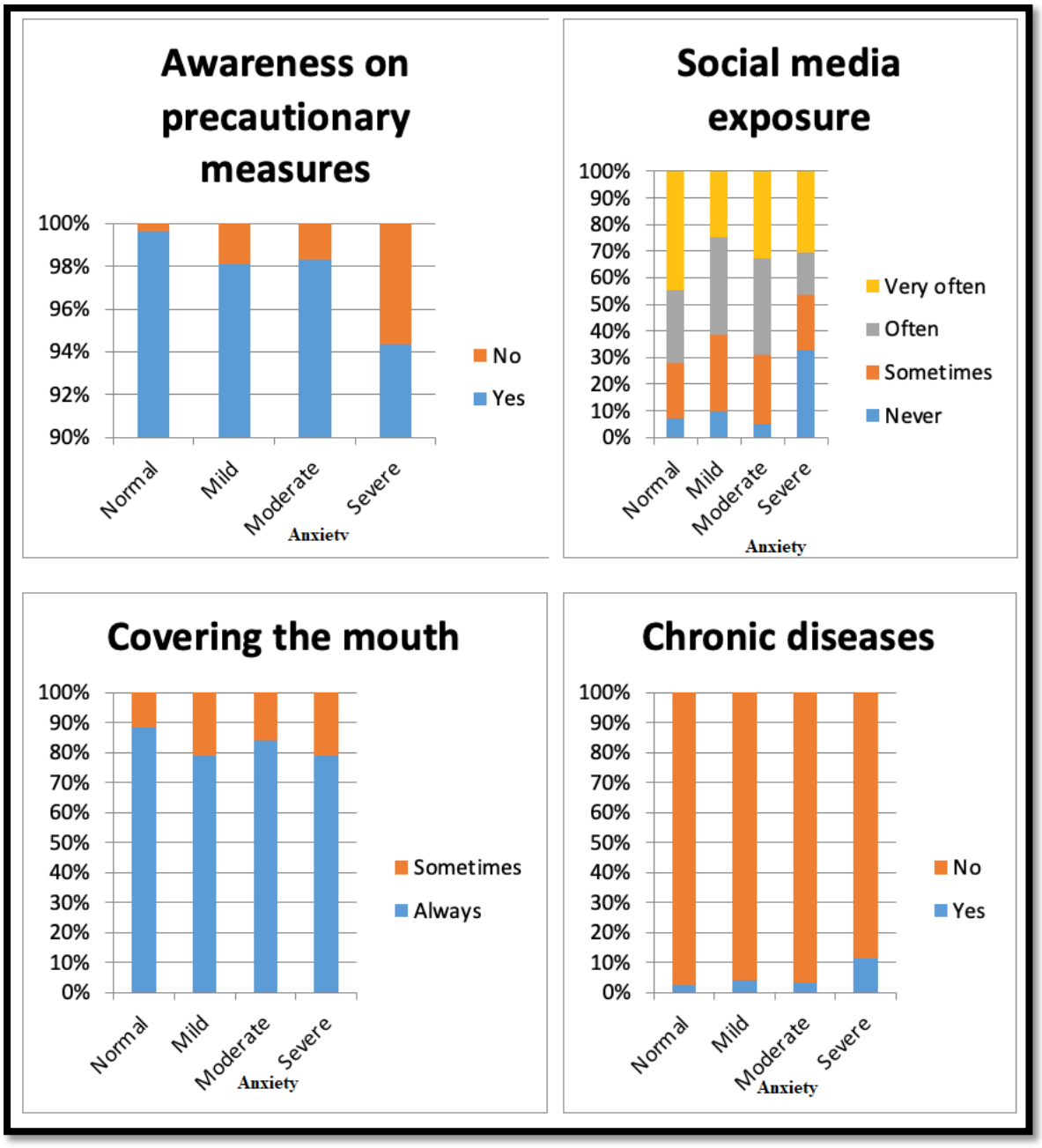


Figure 154: Significant Variables Associated with Anxiety during Lockdown.

Table 76: Psychological Impact by Anxiety

Scores	Anxiety					Total
IESR		Normal	Mild	Moderate	Severe	
	Normal	369(47.4%)	19(2.4%)	30 (3.9%)	11(1.4%)	429(55.1%)
	Mild	83 (10.7%)	11(1.4%)	28 (3.6%)	13(1.7%)	135(17.3%)
	Moderate	29 (3.7%0	6 (0.8%)	12 (1.5%)	7 (0.9%)	54 (6.9%)
	Severe	56 (7.2%)	16(2.1%)	49 (6.3%)	40(5.1%)	161(20.7%)
	Total	537(68.9%)	52(6.7%)	119(15.3%)	71(9.1%)	779

Table 76 represents the anxiety on different levels of psychological impact during the lockdown period. It is clear that out of total respondents 47.4% have normal psychological impact and no anxiety problems. It also indicates that 7.3% individuals are facing mild to moderate anxiety and psychological impact while 5.1% are in severe condition.

Table 77: Stress level by Socio-economic and Demographic Variables

Variable	Groups	Stress				Sig.
		Normal	Mild	Moderate	Severe	
Gender	Male	207	91	29	9	0.976
	Female	266	124	40	13	
	Total	473	215	69	22	
Area of residence	Urban	158	71	28	7	0.673
	Rural	315	144	41	15	
	Total	473	215	69	22	
Marital status**	Single	385	189	64	18	0.031*
	Married	88	26	5	4	
	Total	473	215	69	22	
Educational Qualification	Plus, two or below	34	17	6	1	0.848
	Graduate	234	104	39	10	
	Post Graduate	205	94	24	11	
	Total	473	215	69	22	

Source of information	Friends	98	24	19	4	0.004*
	Mass media	274	122	37	11	
	Social media	101	69	13	7	
	Total	473	215	69	22	
Social media exposure**	Never	34	16	4	3	0.512
	Sometimes	102	53	20	3	
	Often	135	61	23	4	
	Very often	202	85	22	12	
	Total	473	215	69	22	
Covering mouth while coughing and sneezing	Always	412	183	59	17	0.567
	Sometimes	61	32	10	5	
	Total	473	215	69	22	
Washing hands	Frequently	309	135	44	11	0.499
	Sometimes	164	80	25	11	
	Total	473	215	69	22	
Avoid public gathering	Always	385	152	50	13	0.002*
	Sometimes	88	63	19	9	
	Total	473	215	69	22	
Extra activities**	Doing meditation or yoga	46	26	13	3	0.158
	Using social media	77	52	11	4	
	Doing something I am good at	161	63	20	5	
	Spending time with loved ones	95	30	9	4	
	None	94	44	16	6	
	Total	473	215	69	22	
Chronic diseases**	Yes	13	5	8	1	0.002*
	No	460	210	61	21	
	Total	473	215	69	22	
Monthly family income**	Below 5000	40	31	16	6	0.002*
	5000-20000	150	78	23	3	
	20000-50000	164	64	21	5	
	50000-75000	53	20	5	3	
	75000 and above	66	22	4	5	
	Total	473	215	69	22	

Note: * p value is less than 0.05, ** Monte Carlo simulated value.

From Table 77, the factors having influence on stress during lockdown are identified as different sources of information, monthly family income, marital status, avoiding

public gathering and chronic diseases. The measure of significant variables associated with stress during lockdown in Figure 155.

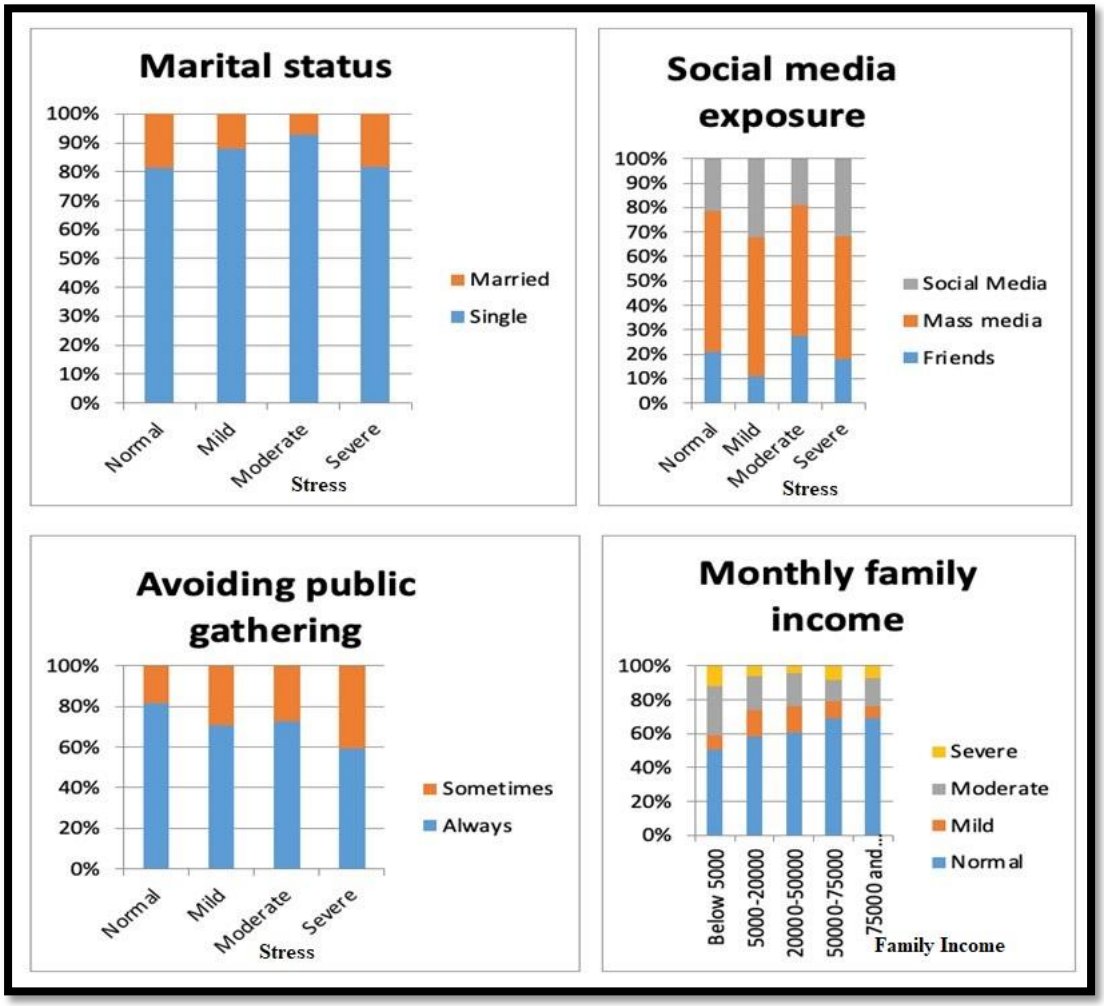


Figure 155: Significant Variables Associated with Stress during Lockdown

Table 78: Baseline Characteristics of Psychological Impact by Stress

Scores	Stress					T
IESR		**N**	**M**	**M**	**S**	
	N	352(45.2%)	66 (8.5%)	11 (1.4%)	0	429 (55.1%)
	M	70 (9%)	53 (6.8%)	11 (1.4%)	1(0.1%)	135 (17.3%)
	M	21 (2.7%)	28 (3.6%)	5 (0.6%)	0	54 (6.9%)
	S	30 (3.9%)	68 (8.7%)	42 (5.4%)	21 (2.7%)	161 (20.7%)
	T	473(60.7%)	215(27.6%)	69 (8.9%)	22 (2.8%)	779

Note: N- Normal, M- Mild, M- Moderate, S- Severe, T- Total

Table 78 represents the stress on different levels of psychological impact during the lockdown period. Here 3.9% of the total respondents have normal stress and severe psychological impact. While, 14.1% of the people are suffering severe psychological impact and mild to moderate stress, whereas, the psychological impact and stress for 2.7% is severe.

Table 79: Depression by Socio-economic and Demographic Variables

Variable	Groups	Depression				Sig.
		Normal	Mild	Moderate	Severe	
Gender	Male	207	35	72	22	0.307
	Female	266	65	82	30	
	Total	473	100	154	52	
Age**	18-25	349	81	126	43	0.188
	26-40	100	15	26	7	
	Above 40	24	4	2	2	
	Total	473	100	154	52	
Area of residence	Urban	151	36	57	20	0.538
	Rural	322	64	97	32	
	Total	473	100	154	52	
Marital status	Single	381	87	143	45	0.003*
	Married	92	13	11	7	
	Total	473	100	154	52	
Educational Qualification**	Plus, two or below	35	9	10	4	0.976
	Graduate	239	49	73	26	
	Postgraduate	199	42	71	22	
	Total	473	100	154	52	
Occupational Status**	Student	253	54	95	31	0.315
	Unemployed	63	13	19	9	
	Healthcare professional	21	8	5	1	
	Gov. employee	25	3	1	2	
	Private employee	111	22	34	9	
	Total	473	100	154	52	
Awareness on precautionary measures**	Yes	470	99	152	49	0.012*
	No	3	1	2	3	
	Total	473	100	154	52	
Source of information	Friends	95	13	21	16	0.006*
	Mass media	279	56	89	44	
	Social media	99	31	44	16	
	Total	473	100	154	52	
Social media exposure**	Never	42	3	9	3	0.298
	Sometimes	101	29	37	11	
	Often	131	33	47	12	
	Very often	199	35	61	26	
	Total	473	100	154	52	

Covering mouth	Always	419	83	127	42	0.103
	Sometimes	54	17	27	10	
	Total	473	100	154	52	
Washing hands	Frequently	321	62	85	31	0.031*
	Sometimes	152	38	69	21	
	Total	473	100	154	52	
Wearing mask**	Yes	450	98	137	47	0.007*
	No	23	2	17	5	
	Total	473	100	154	52	
Avoid public gathering	Always	376	72	115	37	0.208
	Sometimes	97	28	39	15	
	Total	473	100	154	52	
Extra activities	Doing meditation or yoga	50	10	22	6	0.009*
	Using social media	70	23	41	10	
	Doing something I am good at	158	37	40	14	
	Spending time with loved ones	98	16	18	6	
	None	97	14	33	16	
	Total	473	100	154	52	
Chronic diseases**	Yes	10	4	8	5	0.019*
	No	463	96	146	47	
	Total	473	100	154	52	
Monthly family income	Below 5000	47	8	27	11	0.024*
	5000-20000	148	39	51	16	
	20000-50000	155	38	50	11	
	50000-75000	56	8	10	7	
	75000 and above	67	7	16	7	
	Total	473	100	154	52	

* p value is less than 0.05, ** Monte Carlo simulated value.

From Table 79 it is clear that the variables such as marital status, awareness about precautionary measures, washing hands, wearing mask, doing extra activities, monthly family income and chronic diseases are important factors for depression during lockdown.

Table 80: Baseline Characteristics of Psychological Impact by Depression

Score	Depression					Total
IESR		**Normal**	**Mild**	**Moderate**	**Severe**	
	Normal	343 (44%)	35 (4.5%)	45 (5.8%)	6(0.8%)	429 (55.1)
	Mild	66 (8.5%)	22 (2.8%)	41 (5.3%)	6 (0.8%)	135(17.3%)
	Moderate	17 (2.2%)	18 (2.3%)	17 (2.2%)	2 (0.3%)	54 (6.9%)
	Severe	47 (6%)	25 (3.2%)	51 (6.5%)	38(4.9%)	161(20.7%)
	Total	473(60.7)	100(12.8%)	154(19.8%)	52(6.7%)	779

From table 80, 44% of the total respondents showed a normal psychological impact and depression. While, 4.9% are facing severe psychological impact and depression. The graphical representation is given in Figure 156.

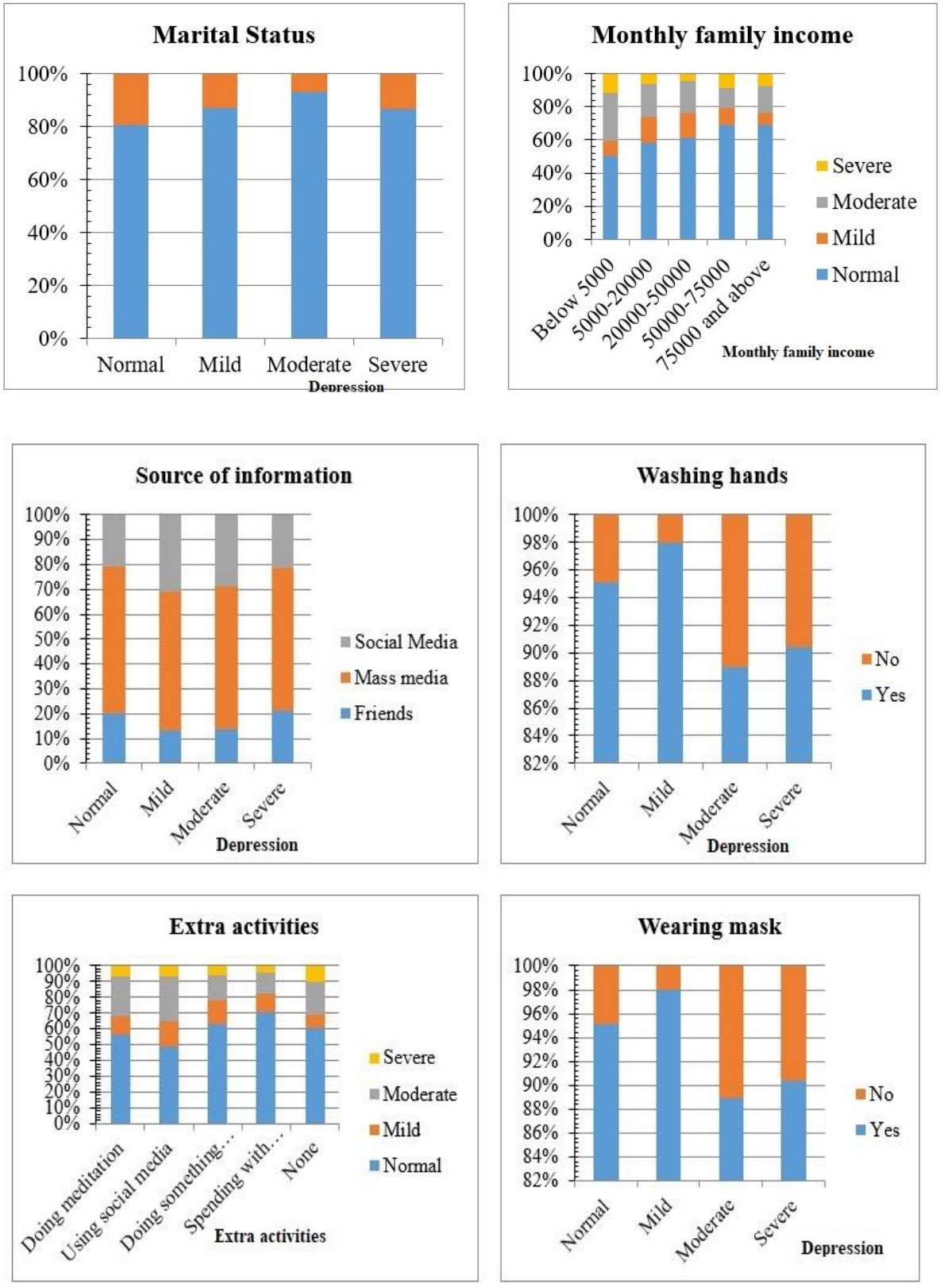


Figure 156: Significant Variables Associated with Depression during Lockdown

Correlation Between IES-R and Depression Anxiety Stress Scale

Based on our data and the scores on the participants' mental health and psychological impact towards COVID-19 there is a significant association between the IES-R, Depression, Anxiety and Stress score (Table 81).

Table 81: Correlation between IES-R and DASS Scale

	Depression	**Anxiety**	**Stress**	**IES-R**
Depression	1.00	0.679*	0.723*	0.592*
Anxiety	0.679*	1.00	0.665*	0.577*
Stress	0.723*	0.665*	1.00	0.684*
IES-R	0.592*	0.577*	0.684*	1.00

Note: * Correlation is significant at 0.05.

Table shows that there is a positive correlation between Depression, Anxiety, Stress and IES-R scales.

Difference in the mean scores in different categories of demographic variables

The normality assumption of the data is tested using Shapiro Wilk test. The null hypothesis to be tested is the data follows normal distribution. Here the null hypothesis is rejected. i.e., normality assumption of the data is violated. So, non parametric tests are used.

Table 82: Difference in Mean Scale Scores for Demographic Variables Having Two Categories

Variable	IESR		Anxiety		Depression		Stress	
	Sig.	Sample Estimate	Sig.	Sample Estimate	Sig.	Sample Estimate	Sig.	Sample Estimate
Gender	<0.001*	20.9	<0.001*	2.99	<0.001*	5.00	<0.001*	7.00
Marital status	<0.001*	21.0	<0.001*	3.00	<0.001*	5.99	<0.001*	7.99
Awareness on precautionary measures	<0.001*	21.0	<0.001*	3.99	<0.001*	5.99	<0.001*	7.99
Chronic diseases	<0.001*	20.0	<0.001*	2.99	<0.001*	4.99	<0.001*	6.99
Area of residence	<0.001*	20.9	<0.001*	2.99	<0.001*	5.00	<0.001*	7.00

Note: * Significant at 0.05

Table 83: Difference in Mean Scale Scores for Demographic Variables having More Than Two Categories

Variables	IESR		Depression		Anxiety		Stress	
	X^2 value	Sig.	X^2 value	Sig.	X^2 value	Sig.	X^2 value	Sig.
Occupational status	62.90	0.65	14.36	0.85	19.10	0.38	18.42	0.62
Educational qualification	55.79	0.85	14.18	0.86	13.68	0.74	15.97	0.77
Social media exposure	94.12	0.01*	23.75	0.30	28.99	0.04*	25.43	0.22
Source of information	81.96	0.11	36.50	0.01*	27.47	0.07	38.21	0.01*
Extra activities	66.56	0.52	30.82	0.07	21.83	0.23	18.38	0.62
Monthly family income	99.02	0.008*	30.38	0.08	35.47	0.008*	38.45	0.01*

Note: * p value is less than 0.05.

Mann Whitney test evident that there is a difference in the mean IES-R score between the two categories of gender, marital status, area of residence, awareness on precautionary measures and chronic diseases (Table 82). There is also a difference in the mean Depression score between the two categories of gender, marital status, area of residence, awareness on precautionary measures and chronic diseases. Similarly, the mean scores of Stress and anxiety shows some difference between the two

categories of gender, marital status, area of residence, awareness on precautionary measures and chronic diseases. The results from the Kruskal Wallis test shows that there is a difference in the mean IES-R score between the various categories of social media exposure and monthly family income. Also, there is a difference in the mean Anxiety score between the various categories of social media exposure and monthly family income. The mean Stress score between the various categories of source of information and monthly family income is also found to be different (Table 83).

Relationship of Variables with IES-R and Depression Anxiety Stress Scales

The normality and homoscedasticity assumptions of the data are tested using Shapiro Wilk test and Breusch-Pagan test. Durbin Watson test is used to check the autocorrelation. The linearity of the data is checked using QQ plot. Since the assumptions are violated the Median regression is used. The variables which are found to be significant in the univariate regression are taken into the multiple regression.

Table 84: Median Regression of Anxiety with Variables Which Are Significant in Univariate Analysis

	Coefficient	Sig.	95% CI	
			LL	UL
Intercept	5.00	0.00	-1.79	5.00
Social media exposure	-1.00	0.001*	-1.00	1.79
Awareness on precautionary m easure	11.00	0.00*	5.25	15.7

Note: * Significant at 0.05, LL-Lower Limit, UL- Upper Limit

Table 84 shows that Anxiety during lockdown period is related with social media exposure and the awareness on precautionary measures. There is a decrease in anxiety

for those people who are exposed to the social media frequently. And the awareness about precautionary measures also input anxiety among people during lockdown.

Table 85: Median Regression of Depression with Variables Which Are Significant in Univariate Analysis

	Coefficient	Significance	95% CI	
			LL	UL
Intercept	14.00	0.00	8.61	17.4
Monthly family income	-0.64	0.064	-1.17	0.12
Marital status	-2.7	0.00*	-4.58	-1.46
Awareness on precautionary measures	8.21	0.15	2.65	16.4
Washing hands	2.85	0.00*	1.34	4.39
Chronic diseases	-7.64	0.002*	-10.61	-2.62
Extra activities	-0.57	0.07	-1.14	0.16
Source of information	0.92	0.13	-0.43	2.52

Note: * Significant at 0.05 LL-Lower Limit, UL- Upper Limit

From table 85, it is seen that there is a decrease of depression among the unmarried people compared to those who are married. The results show that the presence of chronic diseases is also related to depression. It also indicates that the practice of washing hands during this period results in the rise of depression among people.

Table 86: Median Regression of Stress with Variables Which Are Significant in Univariate Analysis

	Coefficient	Sig.	95% CI	
			LL	UL
Intercept	8.00	0.00	4.84	12.12
Monthly family income	-1.33	0.001*	-2.84	0.35
Source of information	2.66	0.02*	-0.54	7.27
Avoid public gathering	1.33	0.05	0.269	3.80

Note: * Significant at 0.05 LL-Lower Limit, UL- Upper Limit

Table 86 shows that Stress during lockdown is related with monthly family income and the source through which people gained information on precautionary measures. There is a rise in stress for those having low monthly family income during lockdown. Similarly, different sources through which people gained information about the COVID-19 outbreak is also an important factor. Mass media is found to be the major source of information and the results suggest that mass media has an influence on the rise of stress among the people during this pandemic.

Table 87: Median Regression of Psychological Impact with Variables Which Are Significant in Univariate Analysis

	Coefficient	**Sig.**	**95% CI**	
			LL	**UL**
Intercept	23.00	0.00	14.06	31.06
Extra activities	4.50	0.001*	1.52	7.72
Avoid public gathering	-1.25	0.01*	-4.62	0.42

Note: * Significant at 0.05 LL-Lower Limit, UL- Upper Limit

Table 87 shows that for the 0.5^{th} quantile the psychological impact of people during lockdown period is related to their extra activities like meditation, yoga and use of social media. Also, there is a decrease in the effect of psychological impact for those people who avoid public gathering during this pandemic.

CHAPTER 4: DISCUSSION

The virus has posed multifaceted challenges from the detection in late 2019, and overwhelming healthcare systems to prompting widespread social and economic disruptions, underscoring the urgent need for comprehensive understanding and effective management strategies.[320] Use of data-driven methods such as consolidated time-series, clustering, and time-series theory for real-time monitoring of COVID-19 helps shape both immediate and strategic long-term responses.[321] These approaches provide helpful insights for timely decision-making and effective management of the pandemic's progression.

Despite the strengths of existing models, their limitations arise from historical data reliance, posing challenges in adapting to the dynamic nature of the pandemic. Data collection process from diverse sources, including open-source COVID-19 databases and supplementary data repositories, ensuring the acquisition of an inclusive dataset that forms the foundation for analyses.[322] With the data curation process undergone, incorporating pre-processing, reconciliation, and fusion techniques to enhance the quality and reliability of the gathered data, ensuring the dataset is robust and reflective of the real-world scenario.

To monitor and evaluate the evolving situation, sophisticated techniques are employed, including consolidated time-series analysis, and model selection criteria.[323] These monitoring techniques contribute to a holistic assessment of the pandemic dynamics, allowing for a nuanced understanding of its various facets. The modelling approaches adopted in this study include SIR (Susceptible-Infectious-Recovered) and exponential growth models, serving as essential tools for estimation and prediction. Making decisions including optimal allocation of limited recourses and impact

assessment provided robust insights, enhancing the ability to make informed, evidence-based decisions for both immediate responses and long-term strategic planning.

Zrieq et al. evaluated the ARIMA statistical technique and the Prophet Facebook machine-learning technique in predicting COVID-19 cases, demonstrating their accuracy and robustness. This study's findings are similar to the findings of Zrieq et al. which highlights the effectiveness of time-series analysis in anticipating surges in cases and resource allocation.[324] A notable difference is the forecasting approach, with Zrieq et al. focusing on out-of-sample forecasting and achieving high coefficients of determination, while the current study employs in-sample forecasting but still demonstrates reasonable accuracy. Both studies underscore the importance of such forecasting models in providing crucial insights for decision-makers and healthcare professionals to prepare for the future evolution of the pandemic and allocate resources effectively.

In the current study, the interrupted time series analysis demonstrated a statistically significant positive coefficient of lockdowns, affirming their effectiveness in altering the trajectory of the COVID-19 pandemic. This finding underscores the importance of implementing lockdown policies as a valuable tool in managing the spread of the virus and reducing its incidence. Conversely, Thayer et al.'s study emphasized the substantial impact of lockdown policies on the pandemic's trajectory. Their interrupted time series analysis revealed an 8% reduction in the change in incidence rate per day after Lockdown 1.0, followed by an additional 3% reduction after Lockdown 4.0, resulting in an overall 11% reduction in COVID-19 incidence.

This study's outcomes further support the significance of lockdown measures in curbing the spread of the virus and reducing its impact on public health.[325]

In the study conducted by Rizvi et al., a comprehensive analysis was carried out across 79 countries, utilizing 18 diverse feature variables encompassing social, economic, health, and environmental metrics. This extensive approach led to the categorization of countries into four clusters, shedding light on the multifaceted factors influencing the spread of COVID-19 on a global scale. The study's global perspective provided valuable insights into the interrelationships among various factors, extending beyond COVID-19 statistics and offering a holistic view that can inform policymakers in their efforts to control the pandemic.[326]

In comparison, the current study focused its clustering efforts on specific states within a single country, India, using COVID-19-specific data such as confirmed cases, deaths, and recovery rates. By concentrating on localized data, this study aimed to identify high-concentration areas within India, enabling targeted resource allocation and attention where it is most needed. Both studies (Rizvi et al., the current study) demonstrate the significance of clustering methods in distilling complex data into meaningful patterns, empowering authorities to tailor their responses, allocate resources efficiently, and implement precise interventions to effectively combat the COVID-19 pandemic, whether at a global or regional level.

In the current study, a comprehensive assessment was conducted using clustering techniques to analyze the shifts in case numbers and recovery rates among various states in India, with a particular focus on the post-lockdown period. The study highlighted the dynamic nature of these clusters and their significance in adapting strategies in real-time, enabling a precise and targeted approach to pandemic control

and economic mitigation. By emphasizing the evolving situation and the effectiveness of clustering in managing COVID-19 metrics across states, the current study offered valuable insights into the ongoing challenges of pandemic management.

On the flip side, Ansari et al.'s analysis concentrated on regional variations in COVID-19 metrics, specifically Case Fatality Rate (CFR) and Recovery Rate (RR) in specific states on specific dates. They pinpointed the states with the highest CFR and the promising recovery rates during specific months in 2020.[327] While Ansari et al. provided a snapshot of certain states during particular timeframes, the current study took a broader perspective, examining the continuous changes and adaptations required in the pandemic control strategies, underscoring the significance of clustering in effectively managing COVID-19 metrics across states.

In the study conducted by Rajendrakumar et al, researchers aimed to estimate the basic reproduction number (R_0) of SARS-CoV-2 in India to understand the virus's transmissibility. The study used crowdsourced time series data and various methods, yielding R_0 estimates ranging from 1.43 to 1.85. The study also modified the SIR model to SIRD to incorporate deaths. According to the findings, the daily epidemic growth rate in India was 0.16 with a doubling time of 4.30 days. The study predicted a peak in mid-July to early August 2020 with approximately 12.5% of the population likely to be infected at that time.[328]

Whereas in the current study, R_0 values of 1.012 and 1.1129 for the SIR and SIRD models, respectively. Although both studies emphasize the significance of R_0 in assessing the virus's spread, the estimated R_0 values differ, possibly due to variations in data sources, methodologies, and periods analyzed. These variations in findings underscore the complexity of modelling infectious disease dynamics and the

importance of considering multiple factors in assessing virus transmissibility and planning response measures.

In the study conducted by Laxminarayan et al., in India provides valuable insights into the dynamics of the COVID-19 pandemic, focusing on different aspects and methodologies. Laxminarayan et al. conducted a prospective, active surveillance study in Madurai, Tamil Nadu, involving a large population and a comprehensive approach to testing and data collection. The study revealed the prevalence of SARS-CoV-2 infections, risk factors associated with symptomatic and asymptomatic cases, and mortality rates across different age groups and comorbidity profiles. The study also assessed seroprevalence, highlighting the discrepancy between clinically diagnosed cases and the actual number of infections. In contrast, the current study focused on state-level data analysis to understand regional variations in the average duration of COVID-19 among patients in different Indian states. This information is essential for tailoring region-specific strategies and resource allocation, as it indicates varying levels of disease severity and healthcare capacity across regions.[329]

In the current study, the significance of predictive modeling in guiding the health sector's response to the COVID-19 pandemic was underscored, particularly through the utilization of the Exponential, Logistic, Gompertz, and Bertalanffy models. Among these models, the Bertalanffy model emerged as the most effective, boasting an impressively high R^2 value of 0.999. This level of accuracy was instrumental for healthcare planners and policymakers, providing them with a reliable foundation for predicting future case numbers and trends. Precise predictions empowered the health sector to allocate resources efficiently, plan hospital capacities meticulously, and devise strategies for medical supplies and staff deployment. This

foresight played a pivotal role in preventing healthcare systems from becoming overwhelmed during the peak periods of the pandemic.

In contrast, the study conducted by Gois et al. shared a similar objective of comprehending COVID-19's behavior and forecasting the total number of infected individuals. Their approach involved a combination of machine learning techniques, logistic regression, filters, and epidemiologic models. Notably, they integrated data from two countries to create a hybrid dataset, achieving promising results with R^2 scores of 0.99 for short-term predictions and 0.93 for long-term predictions. While both studies aimed to predict COVID-19 dynamics, the current study appeared to attain an even higher level of predictive accuracy, particularly with the Bertalanffy model, which could have had a more profound impact on shaping pandemic response planning.[330]

The study conducted by Overton et al. delves into the challenges associated with mathematically modeling and responding to infectious disease outbreaks, particularly focusing on the context of COVID-19. Their work underscores the intricate nature of dealing with data during an outbreak, recognizing the presence of complexities and biases. To address these challenges, they present a toolkit comprising statistical and mathematical models tailored for analyzing early stages of outbreaks and evaluating intervention strategies. This toolkit covers various aspects, including parameter estimation in the presence of data biases and the assessment of non-pharmaceutical interventions within specific subpopulations such as households and care homes.[331] Whereas, the current study takes a distinct approach by emphasizing the utilization of statistical distributions, namely the Weibull, Gamma, and Lognormal distributions, to model COVID-19 mortality data. This approach

offers a robust framework for comprehending and predicting the progression of the pandemic, with a specific focus on mortality trends.

The study identifies the Gamma distribution as the most fitting model, enabling precise predictions of future mortality patterns. This predictive capability is of paramount importance for healthcare authorities as it facilitates effective planning and preparation to tackle the challenges posed by the ongoing pandemic.

Although both studies (Overton et al., current study) aim to provide valuable insights and tools for managing infectious disease outbreaks, they differ in their specific areas of focus. Overton et al. offer a comprehensive toolkit for modeling and assessing interventions, addressing the nuances and biases within outbreak data.[331] Conversely, the current study narrows its scope to mortality data modeling using statistical distributions, underscoring the critical role of accurate predictions in healthcare planning. Combining the approaches from both studies could potentially enhance the understanding and management of infectious disease outbreaks, like COVID-19, by concurrently addressing data complexities and predictive modeling.

The application of the ARIMA model in the current study significantly enhanced the predictive accuracy of COVID-19 death forecasts. The identification of ARIMA (0,1,1) as the most suitable model provided health authorities with a reliable tool to forecast future deaths with a high degree of confidence, as demonstrated in the forecast plot. This critical information played a pivotal role in enabling healthcare decision-makers to anticipate and prepare for evolving healthcare needs, including planning hospital capacities and executing timely public health interventions.

Somyanonthanakul et al. also employed time series analysis techniques to extract valuable insights from COVID-19 data. However, their study primarily focused on predicting COVID-19 hospital admissions and understanding the associated factors, utilizing ARIMA and ARIMAX models.[332] In contrast to the current study's emphasis on mortality predictions, Somyanonthanakul et al. concentrated on hospitalization patterns and patient characteristics. Both studies make valuable contributions to the comprehension of the pandemic's dynamics and offer practical applications for healthcare planning and resource allocation. While one study delves into the intricacies of hospital admissions and patient factors, the other provides crucial insights into mortality predictions, enabling informed public health interventions and capacity planning.

The use of these diverse predictive models provided a multi-faceted understanding of the pandemic's trajectory. This enabled the health sector to respond more effectively to the evolving situation, from managing immediate healthcare needs to planning long-term strategies. The data-driven approach ensured that responses were tailored to the specific needs of different regions, optimizing the use of resources and ultimately contributing to a more effective management of the pandemic. [333]

The data on daily testing rates, adjusted for population size, presented in the current study offered valuable insights into the testing strategies adopted by different countries. This understanding of the scale of testing relative to population size proved instrumental for healthcare workers in assessing the adequacy of testing efforts and gauging the potential presence of undetected cases within communities. Moreover, the utilization of ARIMA models and statistical analysis of COVID-19 data played a significant role in assisting healthcare professionals in effectively managing the

pandemic. These tools provided a data-driven approach for comprehending the virus's spread and severity, facilitating precise and targeted responses, and ultimately enhancing the overall efficiency of the healthcare sector in addressing the crisis.[334]

On the other hand, Malki et al. conducted a study that underscored the importance of daily testing rates adjusted for population size as a means to gain insights into testing strategies employed across various countries. This knowledge is vital for evaluating the sufficiency of testing efforts and identifying potential cases that may have gone undetected. In contrast to the current study's focus on forecasting the slowdown period of the pandemic using a normal distribution model, Malki et al.'s research concentrated on optimizing testing strategies and gaining a deeper understanding of the virus's transmission dynamics.[335] While the current study emphasizes the prediction of infection peaks and rebounds. Together, these studies highlight the crucial role of data-driven insights in addressing the pandemic's challenges and uncertainties and enhancing the effectiveness of healthcare.

The comprehensive analysis of air quality data conducted during the COVID-19 lockdown, utilizing various predictive models such as ARIMA, yielded invaluable insights that greatly benefited the healthcare sector. The observed significant reduction in pollutant levels, including lower concentrations of PM10, PM2.5, NO2, and SO2 across various Indian cities, was directly associated with the decrease in human and industrial activities. This improvement in air quality, evident in the lowered Air Quality Index (AQI) values in cities like Ghaziabad, Delhi, and Kolkata, had a positive impact on public health. The decrease in air pollution levels potentially mitigated respiratory problems, which are particularly relevant during a respiratory disease pandemic like COVID-19. Furthermore, the correlation analysis and

forecasting facilitated by ARIMA models provided a predictive understanding of air quality trends, empowering government monitoring systems to prepare for potential shifts in technology to enhance environmental friendliness and reduce health risks.

Comparably, in a study conducted by Mishra et al., the effects of lockdown and unlock phases on air quality variables across 16 major Indian cities were assessed. This study reported a substantial decrease in AQI values for various pollutants during the lockdown period, with reductions ranging from 30% to 50% for PM2.5, PM10, and CO, and up to 60% for NO2. These findings signify a significant reduction in the atmospheric loading of anthropogenic pollutants due to decreased industrial and vehicular emissions, resulting in improved air quality. The study underscores the substantial impact of pandemic-related measures on air quality in India and highlights the potential for long-term improvements with the implementation of appropriate environmental strategies. These insights were crucial for public health planning and policy-making, particularly in formulating strategies for potential future lockdowns and their implications for both air quality and public health.[336]

The examination of air quality data and its correlation with pandemic-related measures provided valuable insights that contributed to healthcare planning and policy-making. The reduction in pollutant levels during lockdowns had a positive impact on public health, particularly in mitigating respiratory issues during the COVID-19 pandemic. The predictive capabilities of ARIMA models allowed for proactive measures to be taken in anticipation of potential changes in technology and environmental strategies. Additionally, findings from studies like that conducted by Mishra et al. emphasized the substantial improvements in air quality during lockdowns and pointed toward the potential for long-term environmental

enhancements.[336] These insights played a crucial role in shaping public health strategies and preparedness for future pandemic-related measures.

The utilization of time-series analysis and ARIMA modelling for forecasting the growth of active COVID-19 cases in various states of India provided critical insights for healthcare resource allocation during the pandemic. This approach was instrumental in predicting trends and preparing for potential surges in case numbers. The ARIMA (2,2,2) model accurately predicted the fluctuation in active cases, enabling healthcare systems to prepare for changes in ICU, ventilator, and bed demand.[337] The predictive accuracy of this model, as validated by in-sample forecasting, allowed for more informed and strategic resource allocation, ensuring that the healthcare system was not overwhelmed.

Similarly, in Karnataka, the ARIMA (1,2,0) model predicted a steady increase in cases. This foresight was crucial for healthcare planning, allowing authorities to scale up resources in anticipation of rising cases. The ability to forecast demand for ICU beds, ventilators, and hospital beds meant that the state could prepare more effectively for the increasing number of patients. The situation in Maharashtra, predicted using the ARIMA (2,2,0) model, highlighted the state's challenging scenario with a steep increase in active cases. The forecasts enabled healthcare authorities to anticipate the high demand for medical resources and take proactive measures to enhance capacity and manage the crisis more effectively. In another study conducted by Shetty et al., the researchers employed a Multilayer Perceptron (MLP), an Artificial Neural Network (ANN) model, to forecast the number of COVID-19-infected cases in the state of Karnataka, India.[338] To enhance the efficiency of their model, they utilized an Extreme Learning Machine as a fast training algorithm.

The selection of model parameters was performed using both traditional methods like the partial autocorrelation function and the innovative Cuckoo Search (CS) algorithm, a popular metaheuristic optimization approach. The comparison of these parameter selection methods and their MLP model against conventional backpropagation revealed that the use of the CS algorithm resulted in superior forecasting performance, achieving a mean absolute percentage error (MAPE) of 6.62% on training data and 7.03% on test data. To validate the model's robustness, the researchers tested it with COVID-19 data from Hungary, demonstrating a MAPE of 1.55%, thereby establishing the reliability of their ANN model for predicting COVID-19 cases in the state of Karnataka.[338]

In Tamil Nadu, the ARIMA (0,2,1) model's forecasts showed a high growth rate in active cases, indicating an impending strain on healthcare resources. This foresight was crucial in guiding the state's decisions on ramping up healthcare facilities and ensuring adequate availability of ICU beds, ventilators, and general hospital beds. The application of time-series analysis and ARIMA models provided a valuable tool for healthcare authorities in managing the COVID-19 pandemic. By accurately predicting the trend and growth of active cases, states were better equipped to allocate limited healthcare resources effectively, prepare for future demands, and implement strategies to mitigate the impact of the pandemic on their healthcare systems. Similarly, in a study conducted by Kapoor et al, the focus was on addressing the challenges posed by the COVID-19 pandemic in providing healthcare facilities to residents in rural areas with limited access to healthcare services.[339]

The study specifically targeted in a rural area experiencing the highest daily cases, the model projected a need for 57 beds, 8 ICUs, and 2 ventilators per day to meet the healthcare demand.[339] This study demonstrates the utility of the ARIMA model in addressing critical questions related to resource allocation for rural healthcare, offering a versatile tool to enhance healthcare planning and resource distribution in the context of the COVID-19 pandemic.

The social impact of the COVID-19 pandemic on various demographic groups provides valuable insights into the psychological well-being of the population during this challenging period. In the study conducted by Verma et al., the researchers utilized the DASS-21 scale to assess the levels of depression, anxiety, and stress experienced by students during the ongoing COVID-19 pandemic.[340] The study also employed a four-point COPE scale to identify the coping strategies adopted by these students. The results revealed that the pandemic had significantly elevated stress and anxiety levels among the student population.

Although depression levels were not alarmingly high, concerns about various aspects of their lives and future careers were widespread. The findings align with the current study results that indicate while a majority of respondents were aware of precautionary measures, there was still a significant psychological impact, with varying degrees of stress, anxiety, and depression observed among different demographic groups. The high awareness and adherence to precautionary measures like wearing masks and washing hands frequently, combined with a significant reliance on mass media for information, indicate a generally well-informed public. However, this also correlated with increased levels of anxiety and stress, particularly among those who were more exposed to social media.[341] This suggests that while

being informed is crucial, overexposure to pandemic-related news and social media might contribute to psychological distress.

The study's findings offer a comprehensive view of the pandemic's social impact and provide a foundation for healthcare providers to develop targeted mental health interventions. Addressing the psychological impact of the pandemic is as crucial as tackling the physical health challenges, and this research significantly contributes to understanding and meeting these mental health needs.

CHAPTER 5: SUMMARY

5.1 Monitoring

The research begins by addressing the global spread of COVID-19, noting India's lockdown initiation on March 24, 2020, as an effort to curb the pandemic's growth. The research focuses on the effectiveness of the lockdown, evaluating the trends of COVID-19 cases across different Indian states, and deploying time series analysis and hierarchical clustering to study the situation.

In the findings, the ARIMA Model (0, 1, 2) was identified as the best-fitting model for predicting COVID-19 cases in India, with a forecast of 15,097 cases for June 24, 2020. The Interrupted time series analysis demonstrated that the lockdown had a significant impact on COVID-19 cases, as evidenced by a p-value less than 0.05. Hierarchical Cluster Analysis was performed on 15 states with over 500 confirmed cases, categorizing them into three clusters representing varying occurrences of the disease. The study recorded shifts in severity and occurrence across several states but also noted that seven states exhibited no changes.

The research emphasized that caution is required in relaxing quarantine measures, particularly in states with higher occurrences of the disease. Regular hierarchical cluster analysis could inform government decision-making and help in strategizing lockdown relaxation. It was also highlighted that despite the success of the lockdown to some extent, the nature of the virus and various other factors allow for continued growth in India.

The study warns that the pandemic will likely continue for an extended period and that attention should be directed toward new epicentres in the country. Special considerations for protecting medical personnel were deemed essential, as well as

aggressive enhancements in medical facilities' supply. There is also the recommendation to introduce an automated algorithm for regular data fetching and prediction, to aid government and hospitals in planning. The study offers a comprehensive outlook on the lockdown's impact in India, forecasts future trends, and provides strategic recommendations for managing the ongoing COVID-19 crisis.

5.2 Modelling

The analysis identified Tamil Nadu as the most compatible model with an R^2 value of 0.967, while Mizoram exhibited the weakest fit at 0.588. Manipur displayed the shortest average recovery duration (4.94 days), contrasting with Madhya Pradesh's longest duration (68 days). Although data originated from government sources, concerns about data quality, patient follow-up, and various factors such as hospital policies, community fears, healthcare workers' behavior, and diagnostic test accuracy potentially influenced the outcomes. Patient behavior at home and population migrations were recognized as limitations. Given diverse COVID-19 situations and challenges across states, a comprehensive approach considering contextual factors is recommended for meaningful state comparisons.

The distinct characteristics of India's first and second COVID-19 waves, with the latter initiating around February 11, 2021, and varying across states, saw the second phase being considerably more severe, reaching nearly five times the cases of the first wave and peaking at over 0.2 million cases daily on April 14, 2021. India exhibited a diversity of over 7,684 COVID-19 virus variants. By May 2021, the resurgent wave had extended to rural areas, underscoring the need for universal health coverage (UHC). While acknowledging the urgency of saving lives and alleviating suffering, the identified phase characteristics offer insights for managing future

waves. Recognizing the disparate case numbers and healthcare availability at the district level, a tailored, decentralized approach is essential, facilitated by district-level working groups equipped with resources to coordinate responses across all healthcare sectors. Leveraging technology and statistical modelling could streamline resource management for essentials like hospital beds, oxygen, ambulances, and funeral provisions.

The current outlook favours Italy while posing challenges for India and the USA. India needs to escalate testing efforts and disclose the accurate count of affected individuals. It's crucial for India to prepare for worst-case scenarios, focusing on augmenting hospital beds, ventilators, and essential resources. Italy, although making progress, should sustain or accelerate its measures to control the disease, avoiding any reduction in their intensity. Likewise, the USA should adopt a serious approach, persisting with restrictive measures, adhering strictly to regulations such as traffic and travel limitations, bans on gatherings, and the closure of businesses. This collective effort was aided in mitigating the extent of the pandemic's impact.

The global COVID-19 pandemic has emerged as a significant health crisis, presenting India with multifaceted challenges including ensuring access to essential necessities like food, medicine, employment, and shelter. Amid these circumstances, forecasting the disease's future trajectory has gained paramount importance. By calculating an exponential growth rate of 1.127 per day, insights into the disease's expansion have been obtained. The validation of this spread rate through SIR and SIRD models has facilitated a qualitative comprehension of its dissemination. The estimated basic reproduction number derived from epidemiological models underscores the potential for rapid fluctuations as social distancing measures evolve. To address this, we advocate for the continuous enforcement of stringent regulations

to uphold social distancing, ensuring its preservation until the attainment of immunity against this pandemic.

5.3 Making Decision

The analysis suggests a potential healthcare crisis if current trends persist and assumptions hold. Kerala displays the best-case scenario due to its healthcare capacity, while Maharashtra faces a worst-case scenario with rapidly rising cases overwhelming existing facilities. Delhi and Gujarat fall in between. Swift measures such as reinforcing social distancing and reallocating healthcare resources weighted the threat. Employing an SEIR model, the study estimates crucial infection, incubation, and recovery rates. While reliant on assumptions, the study's potential contribution to systematic COVID-19 prevention is significant and emphasizes the importance of proactive measures.

Air quality indices serve as valuable tools to assess ambient air quality, with four methods (AQImean, AQIgm, ORNAQI, and AQIWeiAv) indicating a clear reduction in pollution levels during lockdowns in five selected cities. Notably, average pollutant concentrations of PM10, PM2.5, SO2, and NO2 significantly decreased during this period, leading to a narrower interrelationship between pollutants. Forecasts suggest that pollution rates continue declining until lockdowns end, underscoring how COVID-19 lockdowns improved air quality. These changes highlight the human impact on the environment. To sustain these improvements, necessary pollution control strategies including adopting green protocols, reducing traffic-related pollution, promoting remote work, etc., are crucial. This period of decreased air pollution offers a glimpse of a healthier, cleaner world. Its permanence depends on long-term decisions encompassing politics, institutions, society,

communities, and individual mindsets. Implementing rigorous green protocols in India holds the potential to maintain this positive trend and ensure lasting environmental well-being post-COVID-19.

Infectious disease outbreaks, including the ongoing COVID-19 pandemic, it was evoked fear and impacted mental health. While staying informed is vital, there are strategies to bolster well-being. An assessment of the psychological impact and mental health status of Kerala's general population revealed that half experienced mild to severe psychological impact, with 5.1% facing severe psychological impact and anxiety. The study employed the IES-R and Depression Anxiety Stress scale, finding connections between psychological impact, anxiety, and information sources for precautionary measures. Social media exposure played a role in anxiety development, suggesting the importance of monitoring its content. Strong correlations between IES-R, Depression, Anxiety, and Stress scales were noted. Providing online psychiatric consultation and community-focused efforts by healthcare professionals and government agencies emerged as crucial. Awareness campaigns on precautionary measures played a role in mitigating the negative psychological and mental health effects of the pandemic, fostering resilience and smoother adaptation to challenging circumstances.

CHAPTER 6: CONCLUSION

In-depth analysis of the spread of COVID-19, employing a multifaceted approach that integrated time-series analysis, growth models, and real-time epidemiology to assess the pandemic's trajectory and the effectiveness of lockdown measures. The initial spread of the virus was characterized by an exponential growth pattern, with the application of Exponential, Logistic, Gompertz, and Bertalanffy models proving to be robust in predicting the early trends. The estimation of the basic reproduction number (R_0) indicated that the SIR and SIRD models, along with their modifications, provided accurate predictions in the initial phase of the pandemic within the first 120 days. The second phase had almost five time more when compared with the first and was disastrous, with a number of cases exceeding 0.2 million per day on April 14, 2021.

Lockdown interventions had a significant impact on the trends of COVID-19 cases across various Indian states. Time-series analysis and hierarchical clustering were pivotal in evaluating the effectiveness of these measures. Moreover, the lockdown led to a substantial 37% reduction in air pollutant levels, highlighting the environmental benefits of reduced human activity during this period.

Tamil Nadu was identified as the best fit for the model with an R^2 value of 0.96 and Mizoram was identified to have the least fit with an R^2 value of 0.58. At the same time, Manipur was identified to have the lowest average recovery duration with 4.94 days, and the highest in Madhya Pradesh with 68 days. The analysis also pointed out the potential risks associated with secondary phases of the pandemic, which could peak more quickly and place additional strains on the healthcare system. This

underscores the importance of being prepared for sudden surges in cases and having a responsive healthcare infrastructure.

The socioeconomic implications of the lockdown were also profound. The rise in stress levels among those with lower monthly family incomes, and the significant influence of information sources on public perception and stress, shed light on the psychological and social challenges posed by the pandemic.

The consolidation of time-series data, compartmental models, and impact assessment offers a comprehensive understanding of the COVID-19 pandemic in India. These findings provide valuable insights for policymakers and public health officials in making informed decisions and preparing for future challenges. It also opens avenues for further research on the long-term impacts of the pandemic on various aspects of society and the environment.

6.1 Limitation of Study

1. Statistical models are built upon certain assumptions about the underlying data distribution, relationships, and other factors. These assumptions do not always hold true for the complex and rapidly evolving nature of a pandemic like COVID-19. Deviations from these assumptions led to inaccurate predictions.
2. COVID-19 trends and dynamics have shown significant shifts over time due to various factors like policy changes, public behaviour, new variants, and vaccination efforts. Traditional statistical models struggle to capture these abrupt changes and adapt quickly enough.
3. Statistical models often establish correlations but struggle to determine causal relationships. For example, an increase in reported cases due to increased testing

rather than an actual rise in infections. Distinguishing between correlation and causation is crucial for informed decision-making.

4. Different regions, populations, and demographic groups exhibit varying patterns of COVID-19 transmission and response to interventions. Statistical models that do not adequately account for this heterogeneity yielded generalized and less accurate predictions. The comparison between regions, populations, and demographic groups has its own limitations to a certain extent.
5. Despite the large amount of data collected, there were unobserved or unknown factors that influence the spread of COVID-19. These hidden factors introduced uncertainty and limited the accuracy of statistical models.
6. The use of statistical models for decision-making will have ethical implications, such as resource allocation and prioritization. Models inadvertently perpetuate biases or lead to unfair outcomes if not designed and used thoughtfully.

6.2 Recommendation for Further Study

1. A one-size-fits-all approach is untenable since the numbers of COVID-19 cases and health services differ substantially from district to district in the upcoming waves. Working groups at the district level that have the autonomy to respond to quickly changing local situations must be empowered to receive resources and funds to coordinate efforts across the multifarious sectors of the health care system, i.e. from the front-line workers to tertiary care. Relevant technologies along with statistical modelling could have a role in streamlining the management of resources and commodities, such as oxygen, ambulances, hospital beds, and funeral services.

2. Researchers should explore ways to integrate dynamic factors such as policy changes, vaccination campaigns, and human behaviour into the models. This could involve developing more adaptive models that will quickly adjust to new developments in the pandemic.
3. Explore the integration of diverse data sources, including medical records, mobility data, and social media trends, to enhance the accuracy and comprehensiveness of the models.
4. Develop methods for real-time model updating and recalibration as new data becomes available. This will ensure that the models remain accurate and relevant as the pandemic evolves.
5. Investigate effective ways to communicate model outputs and predictions to the public, policymakers, and healthcare professionals. Ensuring clear and transparent communication is essential for building trust and promoting appropriate actions.

By addressing these recommendations in further studies, the field will advance its understanding of using statistical models for pandemic estimation and monitoring, leading to more effective strategies for managing and controlling future health crises.

6.3 Policy Implications

The study on the use of statistical models for the estimation and monitoring of the COVID-19 pandemic carries significant policy implications for public health authorities, policymakers, and healthcare providers. These implications will guide decisions and actions aimed at managing, containing, and controlling the pandemic effectively:

1. The current study had a very hectic procedure of cleaning and pre-processing data, hence policymakers could prioritize the establishment of standardized data collection, reporting, and sharing mechanisms. Transparent data practices will enhance the accuracy of statistical models and provide a solid foundation for informed decision-making.
2. The study underscores the importance of basing policies and interventions on evidence derived from statistical modelling. Policymakers could actively engage with model outputs to inform strategies such as lockdowns, travel restrictions, vaccination campaigns, and resource allocation.
3. The study highlights the potential for models to identify high-risk areas and populations. Policymakers will use this information to tailor interventions, such as testing and healthcare resources, to areas most in need.
4. Effective resource allocation is critical and from the conducted research it is evident that certain states are in need of it at most. Policymakers will use models to forecast healthcare demands and allocate resources, including hospital beds, ventilators, and medical personnel, to areas with the highest projected needs.
5. The study emphasizes the potential of models that can serve as early warning systems for potential outbreaks or resurgence of cases. Policymakers could use these warnings to implement timely interventions and prevent exponential growth.

Incorporating these policy implications into decision-making processes will enhance the effectiveness of strategies aimed at managing and controlling the COVID-19 pandemic. As the situation evolves, continuous collaboration between researchers, policymakers, healthcare providers, and the public remains essential to navigate the challenges posed by the pandemic effectively.

BIBLIOGRAPHY

1. Sohrabi C, Alsafi Z, O'neill N, Khan M, Kerwan A, Al-Jabir A, Iosifidis C, Agha R. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). International journal of surgery. 2020 Apr 1;76:71-6.
2. Zhao D, Yao F, Wang L, Zheng L, Gao Y, Ye J, Guo F, Zhao H, Gao R. A comparative study on the clinical features of coronavirus 2019 (COVID-19) pneumonia with other pneumonias. Clinical infectious diseases. 2020 Jul 28;71(15):756-61.
3. Nghiem LD, Morgan B, Donner E, Short MD. The COVID-19 pandemic: considerations for the waste and wastewater services sector. Case studies in chemical and environmental engineering. 2020 May 1;1:100006.
4. Shang Y, Li H, Zhang R. Effects of pandemic outbreak on economies: evidence from business history context. Frontiers in public health. 2021 Mar 12;9:146.
5. Onyeaka H, Anumudu CK, Al-Sharify ZT, Egele-Godswill E, Mbaegbu P. COVID-19 pandemic: A review of the global lockdown and its far-reaching effects. Science progress. 2021 May;104(2):00368504211019854.
6. Bchetnia M, Girard C, Duchaine C, Laprise C. The outbreak of the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): A review of the current global status. Journal of infection and public health. 2020 Nov 1;13(11):1601-10.
7. Nasir A, Romero-Severson E, Claverie JM. Investigating the concept and origin of viruses. Trends in microbiology. 2020 Dec 1;28(12):959-67.

8. Araujo SB, Braga MP, Brooks DR, Agosta SJ, Hoberg EP, von Hartenthal FW, Boeger WA. Understanding host-switching by ecological fitting. PLoS One. 2015 Oct 2;10(10):e0139225.
9. Burroughs AL. *Risk factors for virus excretion from a bat colony: An observational study of grey-headed flying foxes (Pteropus poliocephalus) at Eastern Park, Geelong* (Doctoral dissertation, PhD dissertation University of Queensland).
10. Weiss RA. The Leeuwenhoek Lecture 2001. Animal origins of human infectious disease. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences. 2001 Jun 29;356(1410):957-77.
11. Wernery U, Kaaden OR. Viral diseases. Avian medicine. 1st ed. Mosby, London, UK. 2000:266-9.
12. Kumar D, Malviya R, Sharma PK. Corona virus: a review of COVID-19. EJMO. 2020 Mar 4;4(1):8-25.
13. Dhama K, Patel SK, Sharun K, Pathak M, Tiwari R, Yatoo MI, Malik YS, Sah R, Rabaan AA, Panwar PK, Singh KP. SARS-CoV-2 jumping the species barrier: Zoonotic lessons from SARS, MERS and recent advances to combat this pandemic virus. Travel medicine and infectious disease. 2020 Sep 1;37:101830.
14. Morens DM, Fauci AS. Emerging pandemic diseases: how we got to COVID-19. Cell. 2020 Sep 3;182(5):1077-92.
15. Johansen MD, Irving A, Montagutelli X, Tate MD, Rudloff I, Nold MF, Hansbro NG, Kim RY, Donovan C, Liu G, Faiz A. Animal and translational models of SARS-CoV-2 infection and COVID-19. Mucosal immunology. 2020 Nov;13(6):877-91.

16. Parrish CR, Holmes EC, Morens DM, Park EC, Burke DS, Calisher CH, Laughlin CA, Saif LJ, Daszak P. Cross-species virus transmission and the emergence of new epidemic diseases. Microbiology and Molecular Biology Reviews. 2008 Sep;72(3):457-70.
17. World Health Organization. WHO-convened global study of origins of SARS-CoV-2: China Part.
18. Wong S, Lau S, Woo P, Yuen KY. Bats as a continuing source of emerging infections in humans. Reviews in medical virology. 2007 Mar;17(2):67-91.
19. Fan Y, Zhao K, Shi ZL, Zhou P. Bat coronaviruses in China. Viruses. 2019 Mar 2;11(3):210.
20. Dhewantara PW, Mamun AA, Zhang WY, Yin WW, Ding F, Guo D, Hu W, Magalhães RJ. Geographical and temporal distribution of the residual clusters of human leptospirosis in China, 2005–2016. Scientific reports. 2018 Nov 9;8(1):16650.
21. Lamers MM, Haagmans BL. SARS-CoV-2 pathogenesis. Nature reviews microbiology. 2022 May;20(5):270-84.
22. Wang Z, Huang G, Huang M, Dai Q, Hu Y, Zhou J, Wei F. Global patterns of phylogenetic diversity and transmission of bat coronavirus. Science China Life Sciences. 2023 Apr;66(4):861-74.
23. Ni W, Yang X, Yang D, Bao J, Li R, Xiao Y, Hou C, Wang H, Liu J, Yang D, Xu Y. Role of angiotensin-converting enzyme 2 (ACE2) in COVID-19. Critical Care. 2020 Dec;24(1):1-0.
24. Setti L, Passarini F, De Gennaro G, Barbieri P, Licen S, Perrone MG, Piazzalunga A, Borelli M, Palmisani J, Di Gilio A, Rizzo E. Potential role of particulate matter in the spreading of COVID-19 in Northern Italy: first

observational study based on initial epidemic diffusion. BMJ open. 2020 Sep 1;10(9):e039338.

25. Lau BT, Pavlichin D, Hooker AC, Almeda A, Shin G, Chen J, Sahoo MK, Huang CH, Pinsky BA, Lee HJ, Ji HP. Profiling SARS-CoV-2 mutation fingerprints that range from the viral pangenome to individual infection quasispecies. Genome medicine. 2021 Dec;13(1):1-23.
26. Banerjee A, Mukherjee S, Maji BK. Manipulation of genes could inhibit SARS-CoV-2 infection that causes COVID-19 pandemics. Experimental Biology and Medicine. 2021 Jul;246(14):1643-9.
27. Garry RF. The evidence remains clear: SARS-CoV-2 emerged via the wildlife trade. Proceedings of the National Academy of Sciences. 2022 Nov 22;119(47):e2214427119.
28. Ryan JM, Nanda S. COVID-19: Social inequalities and human possibilities. Routledge; 2022 Mar 13.
29. McMichael AJ. Environmental and social influences on emerging infectious diseases: past, present and future. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences. 2004 Jul 29;359(1447):1049-58.
30. Behar JA, Liu C, Kotzen K, Tsutsui K, Corino VD, Singh J, Pimentel MA, Warrick P, Zaunseder S, Andreotti F, Sebag D. Remote health diagnosis and monitoring in the time of COVID-19. Physiological measurement. 2020 Nov 2;41(10):10TR01.
31. Zhou L, Wu Z, Li Z, Zhang Y, McGoogan JM, Li Q, Dong X, Ren R, Feng L, Qi X, Xi J. One hundred days of coronavirus disease 2019 prevention and control in China. Clinical Infectious Diseases. 2021 Jan 15;72(2):332-9.

32. Kumar A, Singh R, Kaur J, Pandey S, Sharma V, Thakur L, Sati S, Mani S, Asthana S, Sharma TK, Chaudhuri S. Wuhan to world: the COVID-19 pandemic. Frontiers in cellular and infection microbiology. 2021 Mar 30;11:596201.

33. Lam WK, Zhong NS, Tan WC. Overview on SARS in Asia and the world. Respirology. 2003 Nov;8:S2-5.

34. Doherty PC. Pandemics: What Everyone Needs to Know®. Oxford University Press; 2013 Jul 31.

35. Jamison DT, Wu KB. The East–West divide in response to COVID-19. Engineering. 2021 Jul 1;7(7):936-47.

36. Husain I, Briggs B, Lefebvre C, Cline DM, Stopyra JP, O'Brien MC, Vaithi R, Gilmore S, Countryman C. Fluctuation of public interest in COVID-19 in the United States: retrospective analysis of Google Trends search data. JMIR public health and surveillance. 2020 Jul 17;6(3):e19969.

37. Indirli M, Di Maio V, Martinelli L. An analysis of Italian resiliance during COVID-19 pandemic: first phase from January to June 2020.

38. Khan N, Fahad S, Naushad M, Faisal S. COVID-2019 Current and Past Situation Analysis in the World and Its Impact on World Economy. Available at SSRN 3715051. 2020 Oct 19.

39. Amare M. *IMPACT OF COVID-19 PANDEMIC, NON-PHARMACEUTICAL, AND PHARMACEUTICAL (VACCINATION) RESPONSE POLICIES ON SUSTAINABLE DEVELOPMENT IN ETHIOPIA: MICROMACRO PERSPECTIVE* (Doctoral dissertation, Haramaya University).

40. Benton M, Batalova J, Davidoff-Gore S, Schmidt T. COVID-19 and the State of Global Mobility in 2020. Geneva: International Organization for Migration; 2021 Apr.

41. Zetzsche DA, Consiglio R. One Million or One Hundred Million Casualties?–The Impact of the COVID-19 Crisis on the Least Developed and Developing Countries. University of Luxembourg, Law Working Paper Series. 2020 Sep 21;8.

42. Cruz BS, de Oliveira Dias M. COVID-19: from Outbreak to Pandemic. GSJ. 2020 Mar;8(3):2230-8.

43. Barber RM, Sorensen RJ, Pigott DM, Bisignano C, Carter A, Amlag JO, Collins JK, Abbafati C, Adolph C, Allorant A, Aravkin AY. Estimating global, regional, and national daily and cumulative infections with SARS-CoV-2 through Nov 14, 2021: a statistical analysis. The Lancet. 2022 Jun 25;399(10344):2351-80.

44. Shakya DR, Thapa SB, Kar SK, Sharma V, Uchida N, Ortiz MR, Chapagain G, Poudel CK, Bhattarai PR. COVID-19 across countries: situation and lessons for pandemic control. Journal of BP Koirala Institute of Health Sciences. 2020 Jul 26;3(1):9-27.

45. Guha S. India in the pandemic age. Indian Economic Review. 2020 Nov; 55:13-30.

46. Lone SA, Ahmad A. COVID-19 pandemic–an African perspective. Emerging microbes & infections. 2020 Jan 1;9(1):1300-8.

47. Lu N, Cheng KW, Qamar N, Huang KC, Johnson JA. Weathering COVID-19 storm: Successful control measures of five Asian countries. American journal of infection control. 2020 Jul 1;48(7):851-2.

48. Oran DP, Topol EJ. The proportion of SARS-CoV-2 infections that are asymptomatic: a systematic review. Annals of internal medicine. 2021 May;174(5):655-62.

49. Sanyaolu A, Okorie C, Hosein Z, Patidar R, Desai P, Prakash S, Jaferi U, Mangat J, Marinkovic A. 446. COVID 19 Pandemicity: a global situation report as of June 9, 2020. InOpen Forum Infectious Diseases 2020 Oct 1 (Vol. 7, No. Supplement_1, pp. S291-S291). US: Oxford University Press.

50. Jackson JK, Weiss MA, Schwarzenberg AB, Nelson RM, Sutter KM, Sutherland MD. Global economic effects of COVID-19.

51. Rabbani MR, Bashar A, Nawaz N, Karim S, Ali MA, Rahiman HU, Alam MS. Exploring the role of islamic fintech in combating the aftershocks of COVID-19: The open social innovation of the islamic financial system. Journal of Open Innovation: Technology, Market, and Complexity. 2021 Jun 1;7(2):136.

52. Buheji M, da Costa Cunha K, Beka G, Mavric B, De Souza YL, da Costa Silva SS, Hanafi M, Yein TC. The extent of COVID-19 pandemic socio-economic impact on global poverty. a global integrative multidisciplinary review. American Journal of Economics. 2020 Aug 1;10(4):213-24.

53. Hoo HE, Loh HC, Ch'ng AS, Hoo FK, Looi I. Positive impacts of the COVID-19 pandemic and public health measures on healthcare. Progress In Microbes & Molecular Biology. 2021 Jul 27;4(1).

54. Akinbi A, Forshaw M, Blinkhorn V. Contact tracing apps for the COVID-19 pandemic: a systematic literature review of challenges and future directions for neo-liberal societies. Health Information Science and Systems. 2021 Dec;9:1-5.

55. Sharma A, Ahmad Farouk I, Lal SK. COVID-19: a review on the novel coronavirus disease evolution, transmission, detection, control and prevention. Viruses. 2021 Jan 29;13(2):202.

56. World Health Organization. Managing epidemics: key facts about major deadly diseases. World Health Organization; 2018.

57. Aduhene DT, Osei-Assibey E. Socio-economic impact of COVID-19 on Ghana's economy: challenges and prospects. International Journal of Social Economics. 2021 Mar 16;48(4):543-56.

58. Hale T, Angrist N, Hale AJ, Kira B, Majumdar S, Petherick A, Phillips T, Sridhar D, Thompson RN, Webster S, Zhang Y. Government responses and COVID-19 deaths: Global evidence across multiple pandemic waves. PLoS One. 2021 Jul 9;16(7):e0253116.

59. Chu IY, Alam P, Larson HJ, Lin L. Social consequences of mass quarantine during epidemics: a systematic review with implications for the COVID-19 response. Journal of travel medicine. 2020 Oct;27(7):taaa192.

60. Haldane V, De Foo C, Abdalla SM, Jung AS, Tan M, Wu S, Chua A, Verma M, Shrestha P, Singh S, Perez T. Health systems resilience in managing the COVID-19 pandemic: lessons from 28 countries. Nature Medicine. 2021 Jun;27(6):964-80.

61. Vuong QH, Le TT, La VP, Nguyen HT, Ho MT, Van Khuc Q, Nguyen MH. COVID-19 vaccines production and societal immunization under the serendipity-mindsponge-3D knowledge management theory and conceptual framework. Humanities and Social Sciences Communications. 2022 Dec;9(1).

62. Woolf K, McManus IC, Martin CA, Nellums LB, Guyatt AL, Melbourne C, Bryant L, Gogoi M, Wobi F, Al-Oraibi A, Hassan O. Ethnic differences in

SARS-CoV-2 vaccine hesitancy in United Kingdom healthcare workers: Results from the UK-REACH prospective nationwide cohort study. The Lancet Regional Health–Europe. 2021 Oct 1;9.

63. Byttebier K. COVID-19 Vaccines and Medicines. InCOVID-19 and Capitalism: Success and Failure of the Legal Methods for Dealing with a Pandemic 2022 Apr 23 (pp. 859-1029). Cham: Springer International Publishing.
64. Peleg K, Bodas M, Hertelendy AJ, Kirsch TD. The COVID-19 pandemic challenge to the All-Hazards Approach for disaster planning. International Journal of Disaster Risk Reduction. 2021 Mar 1;55:102103.
65. Decouttere C, De Boeck K, Vandaele N. Advancing sustainable development goals through immunization: a literature review. Globalization and Health. 2021 Aug 26;17(1):95.
66. Lipsitch M, Finelli L, Heffernan RT, Leung GM, Redd; for the 2009 H1N1 Surveillance Group SC. Improving the evidence base for decision making during a pandemic: the example of 2009 influenza A/H1N1. Biosecurity and bioterrorism: biodefense strategy, practice, and science. 2011 Jun 1;9(2):89-115.
67. McLeod KS. *Melting snow: a re-examination of Dr. John Snow, his dot-map, and the 1854 Broad Street cholera outbreak*(Doctoral dissertation, Carleton University).
68. Alamo T, Reina DG, Gata PM, Preciado VM, Giordano G. Data-driven methods for present and future pandemics: Monitoring, modelling and managing. Annual Reviews in Control. 2021 Jan 1;52:448-64.
69. Vickery J, Atkinson P, Lin L, Rubin O, Upshur R, Yeoh EK, Boyer C, Errett NA. Challenges to evidence-informed decision-making in the context of

pandemics: qualitative study of COVID-19 policy advisor perspectives. BMJ global health. 2022 Apr 1;7(4):e008268.

70. Pappas N, Glyptou K. Accommodation decision-making during the COVID-19 pandemic: Complexity insights from Greece. International Journal of Hospitality Management. 2021 Feb 1;93:102767.
71. Anderson W. The model crisis, or how to have critical promiscuity in the time of COVID-19. Social Studies of Science. 2021 Apr;51(2):167-88.
72. Milgroom MG. Epidemiology and SIR Models. InBiology of Infectious Disease: From Molecules to Ecosystems 2023 Nov 26 (pp. 253-268). Cham: Springer International Publishing.
73. Lv Y, Ma C, Li X, Wu M. Big data driven COVID-19 pandemic crisis management: Potential approach for global health. Archives of Medical Science: AMS. 2021;17(3):829.
74. Nyiwul L. Epidemic Control and Resource Allocation: Approaches and Implications for the Management of COVID-19. Studies in Microeconomics. 2021 Dec;9(2):283-305.
75. Chen L, Xu F, Han Z, Tang K, Hui P, Evans J, Li Y. Strategic COVID-19 vaccine distribution can simultaneously elevate social utility and equity. Nature Human Behaviour. 2022 Nov;6(11):1503-14.
76. Chen L, Xu F, Han Z, Tang K, Hui P, Evans J, Li Y. Strategic COVID-19 vaccine distribution can simultaneously elevate social utility and equity. Nature Human Behaviour. 2022 Nov;6(11):1503-14.
77. Streifel RA, Beebe BL, Veil SR, Sellnow TL. Significant choice and crisis decision making: MeritCare's public communication in the Fen–Phen Case. Journal of Business Ethics. 2006 Dec;69:389-97.

78. Ibrahim NK. Epidemiologic surveillance for controlling COVID-19 pandemic: types, challenges and implications. Journal of infection and public health. 2020 Nov 1;13(11):1630-8.

79. Ndiaye M, Oyewobi SS, Abu-Mahfouz AM, Hancke GP, Kurien AM, Djouani K. IoT in the wake of COVID-19: A survey on contributions, challenges and evolution. Ieee Access. 2020 Oct 12;8:186821-39.

80. Madhav N, Oppenheim B, Gallivan M, Mulembakani P, Rubin E, Wolfe N. Pandemics: risks, impacts, and mitigation.

81. Beijen BA, Van Rijswick HF, Anker HT. The importance of monitoring for the effectiveness of environmental directives: a comparison of monitoring obligations in European environmental directives. Utrecht L. Rev.. 2014;10:126.

82. Abouchacra S, Yaman M, Chandrasekhar Nair S. Lessons Learned from the COVID-19 Pandemic. Applied Drug Research, Clinical Trials and Regulatory Affairs: Formerly Applied Clinical Research, Clinical Trials and Regulatory Affairs. 2021 Apr 1;8(1):70-6.

83. Brown OR. Asymptomatic COVID-19; We don't know what we don't know.

84. Goldstein ND, Burstyn I. On the importance of early testing even when imperfect in a pandemic such as COVID-19. Global epidemiology. 2020 Nov;2:100031.

85. Chang V. An ethical framework for big data and smart cities. Technological Forecasting and Social Change. 2021 Apr 1;165:120559.

86. Paul SK, Chowdhury P, Chowdhury MT, Chakrabortty RK, Moktadir MA. Operational challenges during a pandemic: an investigation in the electronics industry. The International Journal of Logistics Management. 2023 Mar 14;34(2):336-62.

87. Prasad R, Sagar SK, Parveen S, Dohare R. Mathematical modeling in perspective of vector-borne viral infections: a review. Beni-Suef University Journal of Basic and Applied Sciences. 2022 Aug 19;11(1):102.

88. Hilfiker L, Ganguly S. A review of pandemics. Integrated risk of pandemic: COVID-19 impacts, resilience and recommendations. 2020:21-60.

89. Maziarz M, Zach M. Agent-based modelling for SARS-CoV-2 epidemic prediction and intervention assessment: A methodological appraisal. Journal of Evaluation in Clinical Practice. 2020 Oct;26(5):1352-60.

90. Villa F, Bagstad KJ, Voigt B, Johnson GW, Portela R, Honzák M, Batker D. A methodology for adaptable and robust ecosystem services assessment. PloS one. 2014 Mar 13;9(3):e91001.

91. Himanen L, Geurts A, Foster AS, Rinke P. Data-driven materials science: status, challenges, and perspectives. Advanced Science. 2019 Nov;6(21):1900808.

92. Dasaklis TK, Pappis CP, Rachaniotis NP. Epidemics control and logistics operations: A review. International Journal of Production Economics. 2012 Oct 1;139(2):393-410.

93. Hu S, Xiong C, Yang M, Younes H, Luo W, Zhang L. A big-data driven approach to analyzing and modeling human mobility trend under non-pharmaceutical interventions during COVID-19 pandemic. Transportation Research Part C: Emerging Technologies. 2021 Mar 1;124:102955.

94. Twaróg B. An Iterative Model of The COVID-19 Virus Spread Based on Population Density-A Case Study of Poland.

95. Kerr CC, Stuart RM, Mistry D, Abeysuriya RG, Rosenfeld K, Hart GR, Núñez RC, Cohen JA, Selvaraj P, Hagedorn B, George L. Covasim: an agent-based

model of COVID-19 dynamics and interventions. PLOS Computational Biology. 2021 Jul 26;17(7):e1009149.

96. Van Damme W, Dahake R, Delamou A, Ingelbeen B, Wouters E, Vanham G, Van De Pas R, Dossou JP, Ir P, Abimbola S, Van der Borght S. The COVID-19 pandemic: diverse contexts; different epidemics—how and why?. BMJ global health. 2020 Jul 1;5(7):e003098.
97. Nikolopoulos K, Punia S, Schäfers A, Tsinopoulos C, Vasilakis C. Forecasting and planning during a pandemic: COVID-19 growth rates, supply chain disruptions, and governmental decisions. European journal of operational research. 2021 Apr 1;290(1):99-115.
98. Fong SJ, Dey N, Chaki J. Artificial intelligence for coronavirus outbreak. Springer.; 2021.
99. Ferreira CM, Sá MJ, Martins JG, Serpa S. The COVID-19 contagion–pandemic dyad: A view from social sciences. Societies. 2020 Oct 6;10(4):77.
100. Wettstein M, Wahl HW, Schlomann A. The impact of the COVID-19 pandemic on trajectories of well-being of middle-aged and older adults: A multidimensional and multidirectional perspective. Journal of happiness studies. 2022 Oct;23(7):3577-604.
101. Kerr CC, Stuart RM, Mistry D, Abeysuriya RG, Rosenfeld K, Hart GR, Núñez RC, Cohen JA, Selvaraj P, Hagedorn B, George L. Covasim: an agent-based model of COVID-19 dynamics and interventions. PLOS Computational Biology. 2021 Jul 26;17(7):e1009149.
102. Desai AN, Kraemer MU, Bhatia S, Cori A, Nouvellet P, Herringer M, Cohn EL, Carrion M, Brownstein JS, Madoff LC, Lassmann B. Real-time epidemic

forecasting: challenges and opportunities. Health security. 2019 Aug 1;17(4):268-75.

103. Ansari Z, Carson NJ, Ackland MJ, Vaughan L, Serraglio A. A public health model of the social determinants of health. Sozial-und Präventivmedizin/Social and Preventive Medicine. 2003 Jul;48:242-51.
104. Sims CA, Goldfeld SM, Sachs JD. Policy analysis with econometric models. Brookings papers on economic activity. 1982 Jan 1;1982(1):107-64.
105. Demertzis K, Taketzis D, Tsiotas D, Magafas L, Iliadis L, Kikiras P. Pandemic analytics by advanced machine learning for improved decision making of COVID-19 crisis. Processes. 2021 Jul 22;9(8):1267.
106. Ghanim Al-Ani B. Statistical modeling of the novel COVID-19 epidemic in Iraq. Epidemiologic Methods. 2021 Jul 22;10(s1):20200025.
107. Bertozzi AL, Franco E, Mohler G, Short MB, Sledge D. The challenges of modeling and forecasting the spread of COVID-19. Proceedings of the National Academy of Sciences. 2020 Jul 21;117(29):16732-8.
108. Mollison D, editor. Epidemic models: their structure and relation to data. Cambridge University Press; 1995 Jul 13.
109. Holmdahl I, Buckee C. Wrong but useful—what COVID-19 epidemiologic models can and cannot tell us. New England Journal of Medicine. 2020 Jul 23;383(4):303-5.
110. Pielke Jr RA, Conant RT. Best practices in prediction for decision-making: lessons from the atmospheric and earth sciences. Ecology. 2003 Jun; 84(6):1351-8.
111. Fong SJ, Li G, Dey N, Crespo RG, Herrera-Viedma E. Composite Monte Carlo decision making under high uncertainty of novel coronavirus epidemic using

hybridized deep learning and fuzzy rule induction. Applied soft computing. 2020 Aug 1;93:106282.

112. Kai D, Goldstein GP, Morgunov A, Nangalia V, Rotkirch A. Universal masking is urgent in the COVID-19 pandemic: SEIR and agent based models, empirical validation, policy recommendations. arXiv preprint arXiv:2004.13553. 2020 Apr 22.
113. Almeida F, Wasim J. The role of data-driven solutions for SMES in responding to COVID-19. International Journal of Innovation and Technology Management. 2023 Feb 21;20(01):2350001.
114. Sarker IH. Smart City Data Science: Towards data-driven smart cities with open research issues. Internet of Things. 2022 Aug 1;19:100528.
115. Li J, Gong X, Wang Z, Chen R, Li T, Zeng D, Li M. Clinical features of familial clustering in patients infected with 2019 novel coronavirus in Wuhan, China. Virus research. 2020 Sep 1;286:198043.
116. Liu Q, Li Z, Ji Y, Martinez L, Zia UH, Javaid A, Lu W, Wang J. Forecasting the seasonality and trend of pulmonary tuberculosis in Jiangsu Province of China using advanced statistical time-series analyses. Infection and drug resistance. 2019 Jul 26:2311-22.
117. Gupta R, Pandey G, Pal SK. Comparative analysis of epidemiological models for COVID-19 pandemic predictions. Biostatistics & epidemiology. 2021 Jan 2;5(1):69-91.
118. Thompson ER, Williams FS, Giacin PA, Drummond S, Brown E, Nalick M, Wang Q, McDonald JR, Carlson AL. Universal masking to control healthcare-associated transmission of severe acute respiratory coronavirus virus 2 (SARS-CoV-2). Infection Control & Hospital Epidemiology. 2022 Mar;43(3):344-50.

119. Lo-Ciganic WH, Donohue JM, Yang Q, Huang JL, Chang CY, Weiss JC, Guo J, Zhang HH, Cochran G, Gordon AJ, Malone DC. Developing and validating a machine-learning algorithm to predict opioid overdose in Medicaid beneficiaries in two US states: a prognostic modelling study. The Lancet Digital Health. 2022 Jun 1;4(6):e455-65.

120. Faiq M, Kumar A, Singh H, Pareek V, Qadri R, Raza K, Kumari C, Narayan R, Kumar P, Kulandhasamy M, Pandey S. COVID-19: A review on molecular basis, pathogenic mechanisms, therapeutic aspects and future projections 2020.

121. Liu X, Wang L, Ma X, Wang J, Wu L. Modeling the effect of age on quantiles of the incubation period distribution of COVID-19. BMC Public Health. 2021 Dec;21:1-0.21

122. Afzal A, Saleel CA, Bhattacharyya S, Satish N, Samuel OD, Badruddin IA. Merits and limitations of mathematical modeling and computational simulations in mitigation of COVID-19 pandemic: A comprehensive review. Archives of Computational Methods in Engineering. 2022 Mar;29(2):1311-37.

123. Sharma A, Sharma G, Singh F. Computational models to study the infectious disease COVID-19: a review. International Journal of Mathematical Modelling and Numerical Optimisation. 2023;13(4):405-41.

124. Oakley C, Pascoe C, Balthazor D, Bennett D, Gautam N, Isaac J, Isherwood P, Matthews T, Murphy N, Oelofse T, Patel J. Assembly Line ICU: what the Long Shops taught us about managing surge capacity for COVID-19. BMJ open quality. 2020 Dec 1;9(4):e001117.

125. Bennett TD, Moffitt RA, Hajagos JG, Amor B, Anand A, Bissell MM, Bradwell KR, Bremer C, Byrd JB, Denham A, DeWitt PE. Clinical characterization and prediction of clinical severity of SARS-CoV-2 infection among US adults using

data from the US National COVID Cohort Collaborative. JAMA network open. 2021 Jul 1;4(7):e2116901-.

126. Galanis G, Hanieh A. Incorporating social determinants of health into modelling of COVID-19 and other infectious diseases: A baseline socio-economic compartmental model. Social science & medicine. 2021 Apr 1;274:113794.
127. Henley SS, Golden RM, Kashner TM. Statistical modeling methods: challenges and strategies. Biostatistics & Epidemiology. 2020 Jan 1;4(1):105-39.
128. Akman O, Chauhan S, Ghosh A, Liesman S, Michael E, Mubayi A, Perlin R, Seshaiyer P, Tripathi JP. The hard lessons and shifting modeling trends of COVID-19 dynamics: multiresolution modeling approach. Bulletin of Mathematical Biology. 2022 Jan;84:1-30.
129. Tidd J. A review of innovation models. Imperial College London. 2006 Jan 14;16:0-17.
130. Yu S, Qing Q, Zhang C, Shehzad A, Oatley G, Xia F. Data-driven decision-making in COVID-19 response: A survey. IEEE Transactions on Computational Social Systems. 2021 May 28;8(4):1016-29.
131. Krishnan RG, Cenci S, Bourouiba L. Mitigating bias in estimating epidemic severity due to heterogeneity of epidemic onset and data aggregation. Annals of Epidemiology. 2022 Jan 1;65:1-4.
132. Wu J, Wang J, Nicholas S, Maitland E, Fan Q. Application of big data technology for COVID-19 prevention and control in China: lessons and recommendations. Journal of medical Internet research. 2020 Oct 9;22(10):e21980.
133. Ling S, Guo D, Rong Y, Huang GQ. Real-time data-driven synchronous reconfiguration of human-centric smart assembly cell line under graduation

intelligent manufacturing system. Journal of Manufacturing Systems. 2022 Oct 1;65:378-90.1

134. Gama J. A survey on learning from data streams: current and future trends. Progress in Artificial Intelligence. 2012 Apr;1:45-55.
135. Gunessee S, Subramanian N. Ambiguity and its coping mechanisms in supply chains lessons from the COVID-19 pandemic and natural disasters. International Journal of Operations & Production Management. 2020 Nov 20;40(7/8): 1201-23.1
136. Elgendy N, Elragal A, Päivärinta T. DECAS: a modern data-driven decision theory for big data and analytics. Journal of Decision Systems. 2022 Oct 2;31(4):337-73.
137. Tabish SA. COVID-19 pandemic: Emerging perspectives and future trends. Journal of public health research. 2020 Jun 4;9(1):jphr-2020.
138. Kuhl E. Data-driven modeling of COVID-19—Lessons learned. Extreme Mechanics Letters. 2020 Oct 1;40:100921.
139. Prajapati S, Swaraj A, Lalwani R, Narwal A, Verma K. Comparison of traditional and hybrid time series models for forecasting COVID-19 cases. arXiv preprint arXiv:2105.03266. 2021 May 5.
140. Nabi KN. Forecasting COVID-19 pandemic: A data-driven analysis. Chaos, Solitons & Fractals. 2020 Oct 1;139:110046.
141. Rai SS, Rai S, Singh NK. Organizational resilience and social-economic sustainability: COVID-19 perspective. Environment, Development and Sustainability. 2021 Aug;23:12006-23.

142. Rai SS, Rai S, Singh NK. Organizational resilience and social-economic sustainability: COVID-19 perspective. Environment, Development and Sustainability. 2021 Aug;23:12006-23.

143. Gans J. Economics in the Age of COVID-19. mit Press; 2020 May 19.

144. Belso-Martínez JA, Mas-Tur A, Sánchez M, Lopez-Sanchez MJ. The COVID-19 response system and collective social service provision. Strategic network dimensions and proximity considerations. Service Business. 2020 Sep;14: 387-411.

145. Mandal M, Jana S, Nandi SK, Khatua A, Adak S, Kar TK. A model based study on the dynamics of COVID-19: Prediction and control. Chaos, Solitons & Fractals. 2020 Jul 1;136:109889.

146. Majeed A, Hwang SO. Data-driven analytics leveraging artificial intelligence in the era of COVID-19: an insightful review of recent developments. Symmetry. 2021 Dec 23;14(1):16.

147. Kazmi H, Munné-Collado Í, Mehmood F, Syed TA, Driesen J. Towards data-driven energy communities: A review of open-source datasets, models and tools. Renewable and Sustainable Energy Reviews. 2021 Sep 1;148:111290.

148. Wang J, Wang Z. Strengths, weaknesses, opportunities and threats (SWOT) analysis of China's prevention and control strategy for the COVID-19 epidemic. International journal of environmental research and public health. 2020 Apr;17(7):2235.

149. Sheng J, Amankwah-Amoah J, Khan Z, Wang X. COVID-19 pandemic in the new era of big data analytics: Methodological innovations and future research directions. British Journal of Management. 2021 Oct;32(4):1164-83.1

150. Li R, Lu W, Yang X, Feng P, Muqimova O, Chen X, Wei G. Prediction of the epidemic of COVID-19 based on quarantined surveillance in China. medRxiv. 2020 Feb 29:2020-02.

151. Wickramasinghe NC, Steele EJ, Gorczynski RM, Temple R, Tokoro G, Kondakov A, Wallis DH, Klyce B, Wickramasinghe10 DT. Predicting the future trajectory of COVID-19.

152. Cohen JP, Morrison P, Dao L, Roth K, Duong TQ, Ghassemi M. COVID-19 image data collection: Prospective predictions are the future. arXiv preprint arXiv:2006.11988. 2020 Jun 22.

153. Chumachenko D, Chumachenko T, Meniailov I, Pyrohov P, Kuzin I, Rodyna R. On-line data processing, simulation and forecasting of the coronavirus disease (COVID-19) propagation in Ukraine based on machine learning approach. InInternational Conference on Data Stream Mining and Processing 2020 Aug 21 (pp. 372-382). Cham: Springer International Publishing.

154. Hamzah FB, Lau C, Nazri H, Ligot DV, Lee G, Tan CL, Shaib MK, Zaidon UH, Abdullah AB, Chung MH. CoronaTracker: worldwide COVID-19 outbreak data analysis and prediction. Bull World Health Organ. 2020 Mar 19;1(32):1-32.

155. Shuja J, Alanazi E, Alasmary W, Alashaikh A. COVID-19 open source data sets: a comprehensive survey. Applied Intelligence. 2021 Mar;51:1296-325.

156. Kirkby K, Bergen N, Baptista A, Schlotheuber A, Hosseinpoor AR. Data Resource Profile: World Health Organization Health Inequality Data Repository. International Journal of Epidemiology. 2023 Jun 15:dyad078.

157. BN J. Geographic Information Systems and COVID-19: The Johns Hopkins University Dashboard.

158. Terris M. The epidemiologic revolution, national health insurance and the role of health departments. American Journal of Public Health. 1976 Dec;66(12):1155-64.

159. Alamo T, Reina DG, Mammarella M, Abella A. Open data resources for fighting COVID-19. arXiv preprint arXiv:2004.06111. 2020 Apr 13.

160. Christensen H, Turner K, Trickey A, Booton RD, Hemani G, Nixon E, Relton C, Danon L, Hickman M, Brooks-Pollock E, Part of the University of Bristol UNCOVER Group. COVID-19 transmission in a university setting: a rapid review of modelling studies. MedRxiv. 2020 Sep 9:2020-09.

161. Rath S, Tripathy A, Tripathy AR. Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model. Diabetes & metabolic syndrome: clinical research & reviews. 2020 Sep 1;14(5):1467-74.

162. Bej S, Wolkenhauer O. The timing of contact restrictions and pro-active testing balances the socio-economic impact of a lockdown with the control of infections. medRxiv. 2020 May 13:2020-05.

163. Reinhart A, Brooks L, Jahja M, Rumack A, Tang J, Agrawal S, Al Saeed W, Arnold T, Basu A, Bien J, Cabrera ÁA. Beyond Cases and Deaths: The Benefits of Auxiliary Data Streams In Tracking the COVID-19 Pandemic: An open repository of real-time COVID-19 indicators. Proceedings of the National Academy of Sciences of the United States of America. 2021 Dec 12;118(51).

164. Blumenthal D, Fowler EJ, Abrams M, Collins SR. COVID-19—implications for the health care system. New England Journal of Medicine. 2020 Oct 8;383(15):1483-8.

165. DATA CURATION

166. Salem DA, Hashim EM. Impact of data pre-processing on COVID-19 diagnosis using machine learning algorithms. International Journal of Intelligent Systems and Applications in Engineering. 2023 Jan 16;11(1s):164-71.

167. Rao JM, Narayan BH. Novel Coronavirus (COVID-19) Prediction using Deep Learning Model with Improved Meta-Heuristic Optimization Approach. In2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT) 2022 Jan 20 (pp. 935-943). IEEE.

168. Maharana K, Mondal S, Nemade B. A review: Data pre-processing and data augmentation techniques. Global Transitions Proceedings. 2022 Jun 1;3(1):91-9.

169. Tacconelli E, Gorska A, Carrara E, Davis RJ, Bonten M, Friedrich AW, Glasner C, Goossens H, Hasenauer J, Abad JM, Peñalvo JL. Challenges of data sharing in European COVID-19 projects: A learning opportunity for advancing pandemic preparedness and response. The Lancet Regional Health–Europe. 2022 Oct 1;21.

170. Ibrahim NK. Epidemiologic surveillance for controlling COVID-19 pandemic: types, challenges and implications. Journal of infection and public health. 2020 Nov 1;13(11):1630-8.

171. Horaira MA. Impact of COVID-19 Pandemic on Tourism Industry: Possible Reconciliation Strategy for Bangladeshi Tourism Industry. International Tourism and Hospitality Journal. 2021;4(4):1-7.

172. Lenert LA, Ilatovskiy AV, Agnew J, Rudisill P, Jacobs J, Weatherston D, Deans Jr KR. Automated production of research data marts from a canonical fast healthcare interoperability resource data repository: applications to COVID-19 research. Journal of the American Medical Informatics Association. 2021 Aug 1;28(8):1605-11.

173. Pei S, Galanti M, Yamana T, Shaman J. Reconciling Diverse Estimates of COVID-19 Infection Rates.

174. Yin AL, Guo WL, Sholle ET, Rajan M, Alshak MN, Choi JJ, Goyal P, Jabri A, Li HA, Pinheiro LC, Wehmeyer GT. Comparing automated vs. manual data collection for COVID-specific medications from electronic health records. International Journal of Medical Informatics. 2022 Jan 1;157:104622.

175. Lin H, Garg S, Hu J, Wang X, Piran MJ, Hossain MS. Privacy-enhanced data fusion for COVID-19 applications in intelligent Internet of medical Things. IEEE Internet of Things Journal. 2020 Oct 22;8(21):15683-93.

176. Ding W, Nayak J, Swapnarekha H, Abraham A, Naik B, Pelusi D. Fusion of intelligent learning for COVID-19: A state-of-the-art review and analysis on real medical data. Neurocomputing. 2021 Oct 7;457:40-66.

177. Abdullah D, Susilo S, Ahmar AS, Rusli R, Hidayat R. The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data. Quality & Quantity. 2022 Jun;56(3):1283-91.

178. Rosillo N, Del-Águila-Mejía J, Rojas-Benedicto A, Guerrero-Vadillo M, Peñuelas M, Mazagatos C, Segú-Tell J, Ramis R, Gómez-Barroso D. Real time surveillance of COVID-19 space and time clusters during the summer 2020 in Spain. BMC Public Health. 2021 Dec;21(1):1-1.

179. Ajayakumar J, Curtis A, Curtis J. A clustering environment for real-time tracking and analysis of COVID-19 case clusters. InProceedings of the 2nd ACM SIGSPATIAL International Workshop on Spatial Computing for Epidemiology (SpatialEpi 2021) 2021 Nov 2 (pp. 1-9).

180. Maleki M, Mahmoudi MR, Wraith D, Pho KH. Time series modelling to forecast the confirmed and recovered cases of COVID-19. Travel medicine and infectious disease. 2020 Sep 1;37:101742.

181. Jiang F, Zhao Z, Shao X. Time series analysis of COVID-19 infection curve: A change-point perspective. Journal of econometrics. 2023 Jan 1;232(1):1-7.

182. Handayani T. Analysis of Business Success at the Triple 3 M Shop during the COVID-19 Pandemic. InProceeding of International Conference on Business, Economics, Social Sciences, and Humanities 2020 (Vol. 3, pp. 16-20).

183. Pranita M, Marpaung SY, Lubis RD, Ardana N. Monitoring and Evaluation Following Health Protocol. International Archives of Medical Sciences and Public Health. 2021 Dec 29;2(2):213-23.

184. Calonge DS, Connor M, Hultberg P, Shah MA, Aguerrebere PM. Contactless higher education: a SWOT analysis of emergency remote teaching and learning during COVID-19. Journal of Educational Studies and Multidisciplinary Approaches. 2022;2(1).

185. Wang J, Wang Z. Strengths, weaknesses, opportunities and threats (SWOT) analysis of China's prevention and control strategy for the COVID-19 epidemic. International journal of environmental research and public health. 2020 Apr;17(7):2235.

186. Longhurst GJ, Stone DM, Dulohery K, Scully D, Campbell T, Smith CF. Strength, weakness, opportunity, threat (SWOT) analysis of the adaptations to anatomical education in the United Kingdom and Republic of Ireland in response to the COVID-19 pandemic. Anatomical sciences education. 2020 May;13(3):301-11.

187. Al-Hashimy HN. The Impact of Corona virus Pandemic on the International and Domestic Economy: Analysis the Strengths and Weaknesses Based on SWOT Analysis. International Journal of Business and Management Invention. 2022;11(10):90-6.

188. Abid A, Jie S. Impact of COVID-19 on agricultural food: A Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis. Food Frontiers. 2021 Dec;2(4):396-406.

189. Katrakazas C, Michelaraki E, Sekadakis M, Ziakopoulos A, Kontaxi A, Yannis G. Identifying the impact of the COVID-19 pandemic on driving behavior using naturalistic driving data and time series forecasting. Journal of safety research. 2021 Sep 1;78:189-202.

190. Scortichini M, Schneider dos Santos R, De'Donato F, De Sario M, Michelozzi P, Davoli M, Masselot P, Sera F, Gasparrini A. Excess mortality during the COVID-19 outbreak in Italy: a two-stage interrupted time-series analysis. International journal of epidemiology. 2020 Dec 1;49(6):1909-17.

191. Abdullah D, Susilo S, Ahmar AS, Rusli R, Hidayat R. The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data. Quality & Quantity. 2022 Jun;56(3):1283-91.

192. Zarikas V, Poulopoulos SG, Gareiou Z, Zervas E. Clustering analysis of countries using the COVID-19 cases dataset. Data in brief. 2020 Aug 1;31:105787.

193. Virgantari F, Faridhan YE. K-Means clustering of COVID-19 cases in Indonesia's provinces. InProceedings of the International Conference on Global Optimization and Its Applications 2020 Nov 21 (pp. 21-22).

194. Goswami S, Chakrabarti A. Quartile clustering: a quartile based technique for generating meaningful clusters. arXiv preprint arXiv:1203.4157. 2012 Mar 19.

195. Malik M, Sharma AK, Singh P. A Proposed Quartile Clustering Algorithm to Detect Outliers for Large Data Sets.

196. de Oliveira LS, Gruetzmacher SB, Teixeira JP. COVID-19 time series prediction. Procedia Computer Science. 2021 Jan 1;181:973-80.

197. Khan FM, Gupta R. ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India. Journal of Safety Science and Resilience. 2020 Sep 1;1(1):12-8.

198. Verma P, Khetan M, Dwivedi S, Dixit S. Forecasting the COVID-19 outbreak: an application of arima and fuzzy time series models.

199. Ding G, Li X, Jiao F, Shen Y. Brief Analysis of the ARIMA model on the COVID-19 in Italy. medRxiv. 2020 Apr 11:2020-04.

200. Çelik A, Ulu Ç. Testing the Price Bubbles in Cryptocurrencies using Sequential Augmented Dickey-Fuller (SADF) Test Procedures: A Comparison for Before and After COVID-19. Scientific Annals of Economics and Business. 2023 Mar 20;70(1):1-5.

201. Bayyurt L, Bayyurt B. Forecasting of COVID-19 cases and deaths using ARIMA models. medrxiv. 2020 Apr 22:2020-04.

202. Badmus NI, Faweya O, Ige SA. Parametric Modeling Approach to COVID-19 Pandemic Data. arXiv preprint arXiv:2109.06254. 2021 Sep 13.

203. Kumar P, Kalita H, Patairiya S, Sharma YD, Nanda C, Rani M, Rahmani J, Bhagavathula AS. Forecasting the dynamics of COVID-19 pandemic in top 15 countries in April 2020: ARIMA model with machine learning approach. MedRxiv. 2020 Mar 31:2020-03.

204. Bischofberger SM, Hiabu M, Mammen E, Nielsen JP. A comparison of in-sample forecasting methods. Computational Statistics & Data Analysis. 2019 Sep 1;137:133-54.

205. Santos Nobre J, da Motta Singer J. Residual analysis for linear mixed models. Biometrical Journal: Journal of Mathematical Methods in Biosciences. 2007 Jun;49(6):863-75.

206. Hanusz Z, TarasiŃska J. Simulation study on improved Shapiro–Wilk tests for normality. Communications in Statistics-Simulation and Computation. 2014 Oct 21;43(9):2093-105.

207. Bagnato L, De Capitani L, Punzo A. A diagram to detect serial dependencies: an application to transport time series. Quality & Quantity. 2017 Mar;51:581-94.

208. Lee RD, Carter LR. Modeling and forecasting US mortality. Journal of the American statistical association. 1992 Sep 1;87(419):659-71.

209. McDowall D, McCleary R, Bartos BJ. Interrupted time series analysis. Oxford University Press; 2019 Sep 16.

210. Ashcroft DM, Li Wan Po A, Williams HC, Griffiths CE. Clinical measures of disease severity and outcome in psoriasis: a critical appraisal of their quality. British Journal of Dermatology. 1999 Aug 1;141(2):185-91.

211. Alimohamadi Y, Tola HH, Abbasi-Ghahramanloo A, Janani M, Sepandi M. Case fatality rate of COVID-19: a systematic review and meta-analysis. Journal of preventive medicine and hygiene. 2021 Jun;62(2):E311.

212. Karadag E. Increase in COVID-19 cases and case-fatality and case-recovery rates in Europe: a cross-temporal meta-analysis. Journal of medical virology. 2020 Sep;92(9):1511-7.

213. Quah P, Li A, Phua J. Mortality rates of patients with COVID-19 in the intensive care unit: a systematic review of the emerging literature. Critical care. 2020 Dec;24:1-4.

214. Mitra A, Pakhare AP, Roy A, Joshi A. Impact of COVID-19 epidemic curtailment strategies in selected Indian states: an analysis by reproduction number and doubling time with incidence modelling. PLoS One. 2020 Sep 16;15(9):e0239026.

215. Kudryashov NA, Chmykhov MA, Vigdorowitsch M. Analytical features of the SIR model and their applications to COVID-19. Applied Mathematical Modelling. 2021 Feb 1;90:466-73.

216. Cooper I, Mondal A, Antonopoulos CG. A SIR model assumption for the spread of COVID-19 in different communities. Chaos, Solitons & Fractals. 2020 Oct 1;139:110057.

217. Sen D, Sen D. Use of a modified SIRD model to analyze COVID-19 data. Industrial & Engineering Chemistry Research. 2021 Feb 2;60(11):4251-60.

218. Caccavo D. Chinese and Italian COVID-19 outbreaks can be correctly described by a modified SIRD model. MedRxiv. 2020 Mar 23:2020-03.2

219. Shringi S, Sharma H, Rathie PN, Bansal JC, Nagar A. Modified SIRD model for COVID-19 spread prediction for northern and southern states of India. Chaos, Solitons & Fractals. 2021 Jul 1;148:111039.

220. Mahanty C, Kumar R, Mishra BK, Hemanth DJ, Gupta D, Khanna A. Prediction of COVID-19 active cases using exponential and non-linear growth models. Expert Systems. 2022 Mar;39(3):e12648.

221. Laxminarayan R, Vinay TG, Kumar KA, Wahl B, Lewnard JA. SARS-CoV-2 infection and mortality during the first epidemic wave in Madurai, south India: a

prospective, active surveillance study. The Lancet Infectious Diseases. 2021 Dec 1;21(12):1665-76.

222. Chaudhari A, Mankar V, Wetal T, Sakhare S, More P. Design and Development of Interactive, Real-Time Dashboard to Understand COVID-19 Situation in Pune. InData-Driven Approach for Bio-medical and Healthcare 2022 Oct 28 (pp. 131-163). Singapore: Springer Nature Singapore.
223. Jain M, Bhati PK, Kataria P, Kumar R. Modelling logistic growth model for COVID-19 pandemic in India. In2020 5th International Conference on Communication and Electronics Systems (ICCES) 2020 Jun 10 (pp. 784-789). IEEE.
224. Ghosh P, Ghosh R, Chakraborty B. COVID-19 in India: statewise analysis and prediction. JMIR public health and surveillance. 2020 Aug 12;6(3):e20341.
225. Tsoularis A, Wallace J. Analysis of logistic growth models. Mathematical biosciences. 2002 Jul 1;179(1):21-55.
226. Kunnen S, Bosma H. 9 A logistic growth model. A Dynamic Systems Approach of Adolescent Development. 2011:117.
227. Chakraborty B, Bhattacharya S, Basu A, Bandyopadhyay S, Bhattacharjee A. Goodness-of-fit testing for the Gompertz growth curve model. Metron. 2014 Apr;72:45-64.
228. Jose A, Salim A, Subhash S, George N. COVID-19 IN INDIA: MODELLING, FORECASTING AND STATE-WISE COMPARISON. medRxiv. 2020 Jun 16:2020-06.
229. Sharma VK, Nigam U. Modeling and forecasting of COVID-19 growth curve in India. Transactions of the Indian National Academy of Engineering. 2020 Dec;5(4):697-710.

230. Dutta I, Basu T, Das A. Spatial analysis of COVID-19 incidence and its determinants using spatial modeling: A study on India. Environmental Challenges. 2021 Aug 1;4:100096.

231. Mate A, Killian JA, Wilder B, Charpignon M, Awasthi A, Tambe M, Majumder MS. Evaluating COVID-19 lockdown policies for India: A preliminary modeling assessment for individual states. Available at SSRN 3575207. 2020 Apr 15.

232. Goenka MK, Afzalpurkar S, Ghoshal UC, Guda N, Reddy N. Impact of COVID-19 on gastrointestinal endoscopy practice in India: a cross-sectional study. Endoscopy international open. 2020 Jul;8(07):E974-9.

233. Miranda LC, Devezas T. On the global time evolution of the COVID-19 pandemic: Logistic modeling. Technological Forecasting and Social Change. 2022 Feb 1;175:121387.

234. Moreau VH. Using the Weibull distribution to model COVID-19 epidemic data. Model Assisted Statistics and Applications. 2021 Jan 1;16(1):5-14.

235. Aslam M, Rao GS, Saleem M, Sherwani RA, Jun CH. Monitoring mortality caused by COVID-19 using gamma-distributed variables based on generalized multiple dependent state sampling. Computational and Mathematical Methods in Medicine. 2021 Apr 22;2021.

236. Valvo PS. A bimodal lognormal distribution model for the prediction of COVID-19 deaths. Applied Sciences. 2020 Nov 28;10(23):8500.

237. Almetwally EM, Alharbi R, Alnagar D, Hafez EH. A new inverted topp-leone distribution: applications to the COVID-19 mortality rate in two different countries. Axioms. 2021 Feb 26;10(1):25.

238. Yan B, Tang X, Liu B, Wang J, Zhou Y, Zheng G, Zou Q, Lu Y, Tu W. An improved method for the fitting and prediction of the number of COVID-19 confirmed cases based on lstm. arXiv preprint arXiv:2005.03446. 2020 May 5.
239. Mansour MM, A. Farsi M, Mohamed SM, Abd Elrazik EM. Modeling the COVID-19 pandemic dynamics in Egypt and Saudi Arabia. Mathematics. 2021 Apr 10;9(8):827.
240. Aldallal R, Gemeay AM, Hussam E, Kilai M. Statistical modeling for COVID 19 infected patient's data in Kingdom of Saudi Arabia. Plos one. 2022 Oct 28;17(10):e0276688.
241. Bantan RA, Shafiq S, Tahir MH, Elhassanein A, Jamal F, Almutiry W, Elgarhy M. Statistical analysis of COVID-19 data: Using a new univariate and bivariate statistical model. Journal of Function Spaces. 2022 Jun 23;2022.
242. Augustin NH, Sauleau EA, Wood SN. On quantile quantile plots for generalized linear models. Computational Statistics & Data Analysis. 2012 Aug 1;56(8):2404-9.
243. Vrieze SI. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). Psychological methods. 2012 Jun;17(2):228.
244. Harvey AC, Shephard N. 10 Structural time series models.
245. Pyle D. Data preparation for data mining. morgan kaufmann; 1999 Mar 22.
246. Hutabarat AN. Spasial Autocorrelation Analysis On Coronavirus Transmission In East Java 2020. Jurnal Sosial dan Sains. 2022 Jan 15;2(1):69-76.
247. Paparoditis E, Politis DN. The asymptotic size and power of the augmented Dickey–Fuller test for a unit root. Econometric Reviews. 2018 Oct 21;37(9):955-73.

248. Hollerbach J, Khalil W, Gautier M. Model identification. Springer handbook of robotics. 2016:113-38.

249. Garetto M, Leonardi E, Torrisi GL. A time-modulated Hawkes process to model the spread of COVID-19 and the impact of countermeasures. Annual reviews in control. 2021 Jan 1;51:551-63.

250. Levesque J, Maybury DW, Shaw RD. A model of COVID-19 propagation based on a gamma subordinated negative binomial branching process. Journal of theoretical biology. 2021 Mar 7;512:110536.

251. Lyakhnova MV, Rudaev GS. COVID-19 SIMULATION USING ARMA MODELS. Modern Science. 2021(2-1):37-43.

252. Requeijo JG, Cordeiro J. Implementation of the statistical process control with autocorrelated data in an automotive manufacturer. International Journal of Industrial and Systems Engineering. 2013 Jan 1;13(3):325-44.5

253. Korolev I. Identification and estimation of the SEIRD epidemic model for COVID-19. Journal of econometrics. 2021 Jan 1;220(1):63-85.

254. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, Albu E, Arshi B, Bellou V, Bonten MM, Dahly DL. Prediction models for diagnosis and prognosis of COVID-19: systematic review and critical appraisal. bmj. 2020 Apr 7;369.

255. Yadav RS. Data analysis of COVID-2019 epidemic using machine learning methods: a case study of India. International Journal of Information Technology. 2020 Dec;12(4):1321-30.

256. Ibrahim RR, Oladipo HO. Forecasting the spread of COVID-19 in Nigeria using Box-Jenkins modeling procedure. medRxiv. 2020 May 8:2020-05.

257. Bayyurt L, Bayyurt B. Forecasting of COVID-19 cases and deaths using ARIMA models. medrxiv. 2020 Apr 22:2020-04.

258. Tolles J, Luong T. Modeling epidemics with compartmental models. Jama. 2020 Jun 23;323(24):2515-6.

259. Huang R, Liu M, Ding Y. Spatial-temporal distribution of COVID-19 in China and its prediction: A data-driven modeling analysis. The Journal of Infection in Developing Countries. 2020 Mar 31;14(03):246-53.

260. Khan FM, Gupta R. ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India. Journal of Safety Science and Resilience. 2020 Sep 1;1(1):12-8.

261. Meek C, Chickering DM, Heckerman D. Autoregressive tree models for time-series analysis. InProceedings of the 2002 SIAM International Conference on Data Mining 2002 Apr 11 (pp. 229-244). Society for Industrial and Applied Mathematics.

262. Luo J. Forecasting COVID-19 pandemic: Unknown unknowns and predictive monitoring. Technological forecasting and social change. 2021 May 1;166:120602.

263. Hashim MJ, Alsuwaidi AR, Khan G. Population risk factors for COVID-19 mortality in 93 countries. Journal of epidemiology and global health. 2020 Sep;10(3):204.

264. Liang D, Shi L, Zhao J, Liu P, Sarnat JA, Gao S, Schwartz J, Liu Y, Ebelt ST, Scovronick N, Chang HH. Urban air pollution may enhance COVID-19 case-fatality and mortality rates in the United States. The Innovation. 2020 Nov 25;1(3).

265. Shehzad K, Sarfraz M, Shah SG. The impact of COVID-19 as a necessary evil on air pollution in India during the lockdown. Environmental Pollution. 2020 Nov 1;266:115080.

266. Vadrevu KP, Eaturu A, Biswas S, Lasko K, Sahu S, Garg JK, Justice C. Spatial and temporal variations of air pollution over 41 cities of India during the COVID-19 lockdown period. Scientific Reports. 2020 Oct 6;10(1):16574.

267. Plaia A, Ruggieri M. Air quality indices: a review. Reviews in Environmental Science and Bio/Technology. 2011 Jun;10:165-79.

268. Shumway RH, Stoffer DS, Stoffer DS. Time series analysis and its applications. New York: springer; 2000 Mar.

269. Naylor TH, Seaks TG, Wichern DW. Box-Jenkins methods: An alternative to econometric models. International Statistical Review/Revue Internationale de Statistique. 1972 Aug 1:123-37.

270. Benvenuto D, Giovanetti M, Vassallo L, Angeletti S, Ciccozzi M. Application of the ARIMA model on the COVID-2019 epidemic dataset. Data in brief. 2020 Apr 1;29:105340.

271. Samson TK, Ogunlaran OM, Raimi MO. A predictive model for confirmed cases of COVID-19 in Nigeria. TK Samson, OM Ogunlaran, OM Raimi (2020). 2020 Aug 28:1-0.

272. Ison D. Statistical procedures for evaluating trends in coronavirus disease-19 cases in the United States. International Journal of Health Sciences. 2020 Sep;14(5):23.

273. Korolev I. Identification and estimation of the SEIRD epidemic model for COVID-19. Journal of econometrics. 2021 Jan 1;220(1):63-85.

274. Liu Y, Mao C, Leiva V, Liu S, Silva Neto WA. Asymmetric autoregressive models: Statistical aspects and a financial application under COVID-19 pandemic. Journal of Applied Statistics. 2022 Apr 4;49(5):1323-47.

275. Irhaif NH, Kareem ZI, Al-joboori SM. Using the moving average and exponentially weighted moving average with COVID 19. Periodicals of Engineering and Natural Sciences. 2021 Aug 11;9(3):625-31.

276. Lyakhnova MV, Rudaev GS. COVID-19 SIMULATION USING ARMA MODELS. Modern Science. 2021(2-1):37-43.

277. Chujai P, Kerdprasop N, Kerdprasop K. Time series analysis of household electric consumption with ARIMA and ARMA models. InProceedings of the international multiconference of engineers and computer scientists 2013 Mar 13 (Vol. 1, pp. 295-300). Hong Kong: IAENG.

278. Theerthagiri P. Probable forecasting of epidemic COVID-19 in using COCUDE model. EAI Endorsed Transactions on Pervasive Health and Technology. 2021 Feb 3;7(26):e3.

279. Ghosh P, Ghosh R, Chakraborty B. COVID-19 in India: statewise analysis and prediction. JMIR public health and surveillance. 2020 Aug 12;6(3):e20341.

280. Ganiny S, Nisar O. Mathematical modeling and a month ahead forecast of the coronavirus disease 2019 (COVID-19) pandemic: an Indian scenario. Modeling earth systems and environment. 2021 Mar;7(1):29-40.

281. Prajapati D, Kanojia M. Forecasting of COVID-19 Cases in INDIA Using ARIMA and AR Time-Series Algorithm. InInternational Conference on Soft Computing and Pattern Recognition 2021 Dec 15 (pp. 361-370). Cham: Springer International Publishing.

282. Katoch R, Sidhu A. An application of ARIMA model to forecast the dynamics of COVID-19 epidemic in India. Global Business Review. 2021:0972150920988653.

283. Dudley R. THE SHAPIRO–WILK TEST FOR NORMALITY.

284. Ibrahim RR, Oladipo HO. Forecasting the spread of COVID-19 in Nigeria using Box-Jenkins modeling procedure. medRxiv. 2020 May 8:2020-05.

285. Perc M, Gorišek Miksić N, Slavinec M, Stožer A. Forecasting COVID-19. Frontiers in physics. 2020 Apr 8;8:127.

286. Shcherbakov MV, Brebels A, Shcherbakova NL, Tyukov AP, Janovsky TA, Kamaev VA. A survey of forecast error measures. World applied sciences journal. 2013 Sep;24(24):171-6.

287. Chordia S, Pawar Y. Analyzing and forecasting COVID-19 outbreak in India. In2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) 2021 Jan 28 (pp. 1059-1066). IEEE.

288. Saif S, Das P, Biswas S. A hybrid model based on mba-anfis for COVID-19 confirmed cases prediction and forecast. Journal of The Institution of Engineers (India): Series B. 2021 Jan 19:1-4.

289. Hedberg EC, Ayers S. The power of a paired t-test with a covariate. Social science research. 2015 Mar 1;50:277-91.

290. Sedgwick P. Pearson's correlation coefficient. Bmj. 2012 Jul 4;345.29

291. Hussain AA, Al-Turjman F. Resource allocation in volunteered cloud computing and battling COVID-19. AI-powered IoT for COVID-19. 2020 Dec 10:39-76.

292. Jana S, Majumder R, Ghose D. Critical medical resource allocation during COVID-19 pandemic. In5th World Congress on Disaster Management: Volume III 2023 Feb 16 (pp. 275-283). Routledge.

293. Tirupakuzhi Vijayaraghavan BK, Nainan Myatra S, Mathew M, Lodh N, Vasishtha Divatia J, Hammond N, Jha V, Venkatesh B. Challenges in the delivery of critical care in India during the COVID-19 pandemic. Journal of the Intensive Care Society. 2021 Nov;22(4):342-8.

294. Gmach D, Rolia J, Cherkasova L, Kemper A. Capacity management and demand prediction for next generation data centers. InIEEE International Conference on Web Services (ICWS 2007) 2007 Jul 9 (pp. 43-50). IEEE.

295. IHME COVID-19 Health Service Utilization Forecasting Team, Murray CJ. Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months. MedRxiv. 2020 Mar 30:2020-03.

296. Nag S, Chakrabarty SP. Modeling the Dynamics Of COVID-19 Transmission In India: Social Distancing, Regional Spread and Healthcare Capacity. Journal of Biological Systems. 2022 Sep 5;30(03):647-72.

297. Mandal S, Bhatnagar T, Arinaminpathy N, Agarwal A, Chowdhury A, Murhekar M, Gangakhedkar RR, Sarkar S. Prudent public health intervention strategies to control the coronavirus disease 2019 transmission in India: A mathematical model-based approach. The Indian journal of medical research. 2020 Feb;151(2-3):190.

298. Vinod A, Kaimal RS. A cross sectional study to assess the psychological impact of COVID 19 on family medicine specialists and residents in Kerala. Journal of Family Medicine and Primary Care. 2022 Sep 1;11(9):5055-9.

299. Nair DR, Rajmohan V, Raghuram TM. Impact of COVID-19 lockdown on lifestyle and psychosocial stress-an online survey. Kerala Journal of Psychiatry. 2020 Jun 21;33(1):5-15.

300. Parkitny L, McAuley J. The depression anxiety stress scale (DASS). Journal of physiotherapy. 2010 Jan 1;56(3):204.

301. Akin A, Çetin B. The Depression Anxiety and Stress Scale (DASS): The study of validity and reliability. Kuram ve Uygulamada Egitim Bilimleri. 2007;7(1):260.

302. Christianson S, Marren J. The impact of event scale-revised (IES-R). Medsurg Nurs. 2012 Sep 1;21(5):321-2.

303. Ugoni A, Walker BF. The Chi square test: an introduction. COMSIG review. 1995 Nov 11;4(3):61.

304. Tallarida RJ, Murray RB, Tallarida RJ, Murray RB. Chi-square test. Manual of pharmacologic calculations: With computer programs. 1987:140-2.

305. McHugh ML. The chi-square test of independence. Biochemia medica. 2013 Jun 15;23(2):143-9.5

306. Turhan NS. Karl Pearson's Chi-Square Tests. Educational Research and Reviews. 2020 Sep;16(9):575-80.

307. Dudley R. THE SHAPIRO–WILK TEST FOR NORMALITY.

308. Nachar N. The Mann-Whitney U: A test for assessing whether two independent samples come from the same distribution. Tutorials in quantitative Methods for Psychology. 2008 Mar 1;4(1):13-20.

309. Inoue J. Statistical analyses were performed using Mann-Whitney U test for comparison of continuous variables between two groups. Differ.

310. Okeh UM. Statistical analysis of the application of Wilcoxon and Mann-Whitney U test in medical research studies. Biotechnology and molecular biology reviews. 2009 Dec;4(6):128-31.
311. MacFarland TW, Yates JM, MacFarland TW, Yates JM. Mann–whitney u test. Introduction to nonparametric statistics for the biological sciences using R. 2016:103-32.
312. McKight PE, Najab J. Kruskal-wallis test. The corsini encyclopedia of psychology. 2010 Jan 30:1-.
313. Xu H, Deng Y. Dependent evidence combination based on shearman coefficient and pearson coefficient. Ieee Access. 2017 Dec 18;6:11634-40.
314. Pearson K. On further methods of determining correlation. Dulau and Company; 1907.
315. Binder A. Considerations of the place of assumptions in correlational analysis. American Psychologist. 1959 Aug;14(8):504.
316. He X, Fu B, Fung WK. Median regression for longitudinal data. Statistics in medicine. 2003 Dec 15;22(23):3655-69.
317. Brown BM. Median estimates in simple linear regression. Australian Journal of Statistics. 1980 Jun;22(2):154-65.
318. Hao L, Naiman DQ. Quantile regression. Sage; 2007 Apr 18.
319. Davino C, Furno M, Vistocco D. Quantile regression: theory and applications. John Wiley & Sons; 2013 Dec 31.
320. Roychoudhury S, Das A, Sengupta P, Dutta S, Roychoudhury S, Choudhury AP, Ahmed AF, Bhattacharjee S, Slama P. Viral pandemics of the last four decades: pathophysiology, health impacts and perspectives. International journal of environmental research and public health. 2020 Dec;17(24):9411.

321. Horton R. The COVID-19 catastrophe: What's gone wrong and how to stop it happening again. John Wiley & Sons; 2021 Jan 14.

322. Mandal M, Jana S, Nandi SK, Khatua A, Adak S, Kar TK. A model based study on the dynamics of COVID-19: Prediction and control. Chaos, Solitons & Fractals. 2020 Jul 1;136:109889.

323. Zeroual A, Harrou F, Dairi A, Sun Y. Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study. Chaos, solitons & fractals. 2020 Nov 1;140:110121.

324. Zrieq R, Boubaker S, Kamel S, Alzain M, Algahtani FD. Analysis and modeling of COVID-19 epidemic dynamics in Saudi Arabia using SIR-PSO and machine learning approaches. The Journal of Infection in Developing Countries. 2022 Jan 31;16(01):90-100.

325. Thayer WM, Hasan MZ, Sankhla P, Gupta S. An interrupted time series analysis of the lockdown policies in India: a national-level analysis of COVID-19 incidence. Health policy and planning. 2021 Jun 1;36(5):620-9.

326. Rizvi SA, Umair M, Cheema MA. Clustering of countries for COVID-19 cases based on disease prevalence, health systems and environmental indicators. Chaos, Solitons & Fractals. 2021 Oct 1;151:111240.

327. Ansari AA, Desai HD, Sharma K, Jadeja DM, Patel R, Patel Y, Desai HM. Prevalence and cross states comparison of case fatality rate and recovery rate of COVID 19/SARS-COV-2 in India. Journal of Family Medicine and Primary Care. 2021 Jan;10(1):475.

328. Rajendrakumar AL, Nair AT, Nangia C, Chourasia PK, Chourasia MK, Syed MG, Nair AS, Nair AB, Koya MS. Epidemic landscape and forecasting of

SARS-CoV-2 in India. Journal of epidemiology and global health. 2021 Mar;11(1):55.

329. Laxminarayan R, Wahl B, Dudala SR, Gopal K, Mohan B C, Neelima S, Jawahar Reddy KS, Radhakrishnan J, Lewnard JA. Epidemiology and transmission dynamics of COVID-19 in two Indian states. Science. 2020 Nov 6;370(6517):691-7.
330. Góis AN, Laureano EE, Santos DD, Sánchez DE, Souza LF, Vieira RD, Oliveira JC, Santana-Santos E. Lockdown as an Intervention Measure to Mitigate the Spread of COVID-19: a modeling study. Revista da Sociedade Brasileira de Medicina Tropical. 2020 Oct 21;53.
331. Overton CE, Stage HB, Ahmad S, Curran-Sebastian J, Dark P, Das R, Fearon E, Felton T, Fyles M, Gent N, Hall I. Using statistics and mathematical modelling to understand infectious disease outbreaks: COVID-19 as an example. Infectious Disease Modelling. 2020 Jan 1;5:409-41.
332. Somyanonthanakul R, Warin K, Amasiri W, Mairiang K, Mingmalairak C, Panichkitkosolkul W, Silanun K, Theeramunkong T, Nitikraipot S, Suebnukarn S. Forecasting COVID-19 cases using time series modeling and association rule mining. BMC Medical Research Methodology. 2022 Nov 1;22(1):281.
333. Ospina R, Gondim JA, Leiva V, Castro C. An overview of forecast analysis with ARIMA models during the COVID-19 pandemic: Methodology and case study in Brazil. Mathematics. 2023 Jul 12;11(14):3069.
334. Abdo AS, Abdul-Kader HM, Salem RK. Enhanced compressed maximal frequent patterns from COVID-19 streaming data. Studies in Informatics and Control. 2022 Mar 1;31(1):99-108.

335. Malki Z, Atlam ES, Ewis A, Dagnew G, Alzighaibi AR, ELmarhomy G, Elhosseini MA, Hassanien AE, Gad I. ARIMA models for predicting the end of COVID-19 pandemic and the risk of second rebound. Neural Computing and Applications. 2021 Apr;33:2929-48.

336. Mishra M, Kulshrestha UC. A brief review on changes in air pollution scenario over South Asia during COVID-19 lockdown. Aerosol and Air Quality Research. 2021 Apr;21(4):200541.

337. Sahoo H, Mandal C, Mishra S, Banerjee S. Burden of COVID-19 pandemic in India: perspectives from health infrastructure. medRxiv. 2020 May 27:2020-05.

338. Shetty P, Mandal R. APPROACHES TO COMPREHENSIVE AND EFFECTIVE REHABILITATION IN THE COVID 19 PANDEMIC: ASSESSMENT OF AN OFT NEGLECTED CRITICAL VARIABLE IN NATIONAL AND GLOBAL HEALTH POLICY MAKING. Turkish Journal of Physiotherapy and Rehabilitation.;32:3.

339. Kapoor G, Hauck S, Sriram A, Joshi J, Schueller E, Frost I, Balasubramanian R, Laxminarayan R, Nandi A. State-wise estimates of current hospital beds, intensive care unit (ICU) beds and ventilators in India: Are we prepared for a surge in COVID-19 hospitalizations?. MedRxiv. 2020 Jun 18:2020-06.

340. Verma S, Mishra A. Depression, anxiety, and stress and socio-demographic correlates among general Indian public during COVID-19. International Journal of Social Psychiatry. 2020 Dec;66(8):756-62.

341. Roy D, Tripathy S, Kar SK, Sharma N, Verma SK, Kaushal V. Study of knowledge, attitude, anxiety & perceived mental healthcare need in Indian population during COVID-19 pandemic. Asian journal of psychiatry. 2020 Jun 1;51:102083.

ANNEXURE - I - PUBLICATIONS

COVID-19 pandemic and its average recovery time in Indian states

Clinical Epidemiology and Global Health 11 (2021) 100740

Contents lists available at ScienceDirect

Clinical Epidemiology and Global Health

journal homepage: www.elsevier.com/locate/cegh

Original article

COVID-19 pandemic and its average recovery time in Indian states

Noel George *, Naresh K. Tyagi, Jang Bahadur Prasad

Department of Epidemiology and Biostatistics, KAHER, Belagavi, 590010, Karnataka, India

ARTICLE INFO

Keywords:
COVID-19
Recovery time
Models
Indian states
Degree polynomials

ABSTRACT

Background: Many studies have been carried out in modelling COVID-19 pandemic. However, region-wise average duration of recovery from COVID-19 has not been attempted; hence, an effort has been made to estimate state-wise recovery duration of India's COVID-19 patients. Determining the recovery time in each region is intended to assist healthcare professionals in providing better care and planning of logistics.

Methods: This study used database provided by Kaggle, which takes data from the Ministry of Health & Family Welfare. The simple Linear Regression model between incidence, prevalence, and duration was used to assess the duration of COVID-19 disease in various Indian states.

Results: The fitted model suits ideal for most of the states, except for some union territories and northeastern states. The average time to recover from disease was ranging from 5 to 36 days in Indian states/union territories except for Madhya Pradesh. Tamil Nadu has an average recovery time of 7 days with an value of 0.96, followed by Odisha, Karnataka, West Bengal, Kerala and Chhattisgarh and the average recovery duration was estimated as 7, 13, 17, 11, 14 and 12 days respectively.

Conclusion: The average recovery from COVID-19 was ten or less days in twenty percentage of states, whereas in forty-four percentage of states/union territories had an average recovery duration between ten to twenty days. However, around twentyfour percentage of states/union territory recovered between twenty to thirty days. In the rest of Indian states/union territories, the average duration of recovery was more than thirty days.

1. Introduction

The COVID-19 was a human tragedy infecting around 99.86 million individuals, resulting in around 2.1 million deaths by January 27, 2021. India recorded the first case of COVID-19 on January 30, 2020, and cases increased steadily to hit about 10.68 million cases and 153,724 deaths by January 27, 2021. Factors like the stage of disease at admission, patient's care and so forth sway the average COVID recovery duration explicitly. The Ministry of Health has released home isolation guidelines (April 7, 2020) for very mild and pre-symptomatic COVID-19 cases to remain at home quarantine.[1] Furthermore, the Government and hospital's policy in admitting patients are expected to significantly influence the actual recovery duration.[7] According to the World Health Organisation, an act of violence related to the COVID-19 pandemic takes place due to the social stigma and discriminatory behaviours against anyone perceived to have been in contact with the virus or affected by the disease. A similar fear in the community of getting admitted to a hospital and the healthcare workers' behaviour towards the COVID-19 patients also act as possible factors in affecting the average duration. As per the reports by The Lancet (2020), some Indian states had not recorded any suspected case, or probable COVID-19 deaths may also alter the recovery duration of COVID cases.[14] The improper follow-ups of patients, sensitivity and specificity of the diagnostic tests are other possible factors affecting the average duration of disease.

The incubation period of COVID-19 has been reported as 14 days with a 95% confidence interval of 8–16 days.[2] If the incidence of a disease is very high, and the recovery time of the disease is also high, then the prevalence of the disease increases, which puts extra health, economic and social burden on the country's economy.[12] A piolet study conducted in India by Barman MP et al. (2020) on 221 COVID patients found that the probability of recovery from COVID-19 in 10 days is minimal; only 4% of the patients recovered in 10 days, and about 40% of the patients in 20 days. In 25 days of treatment, about 50% of the patients were discharged from the hospital. The study also concluded that the average recovery duration of COVID-19 patients in India had been found 25 days, with a 95% confidence interval of 16 days–34 days.[9] The average duration of COVID-19 is generally computed from hospital-based studies rather than from the community. Therefore, the

* Corresponding author.
E-mail addresses: noelgeorge2007@gmail.com (N. George), nareshktyagi@gmail.com (N.K. Tyagi), jbiips12@gmail.com (J.B. Prasad).

https://doi.org/10.1016/j.cegh.2021.100740
Received 8 March 2021; Received in revised form 27 March 2021; Accepted 4 April 2021
Available online 15 April 2021

present study has been planned to determine the average duration of COVID-19 patients from the community.

Many studies have been carried out to model the COVID-19 pandemic in India; however, there are only a few peer-reviewed articles on the recovery period of patients with COVID-19 where the period of a patient in disease has been assessed. Knowledge of the average duration of recovery time helps the Government to prepare effective measures for the healthcare system.[9] Determining the average duration of recovery time from COVID-19 in each region is expected to help the health professionals for better treatment and arrangement of logistics. Hence, an effort has been made to estimate the recovery duration of India's COVID-19 patients for each state which intern expected to establish a social and economic initiative to help COVID-19 patients during the disease period and thereafter.

2. Methods & materials

This study has used an open-source database provided by Kaggle, which takes the data from the Ministry of Health & Family Welfare, Government of India. A statistical model was developed to study the duration of patients in COVID-19 for each state of India. Data was taken directly from the first day of reporting for each state. During data cleaning and checking for duplicates, Telangana was removed from the analysis due to the inappropriateness of the reported data. The complete data of Tripura was also not available in the dataset.

The standard relationship between Incidence & Prevalence[10,11,13]:

$$\text{Prevalence} = \text{Incidence}*\text{Duration} \tag{1}$$

$$Duration = \frac{Prevalence}{Incidence} = \frac{\frac{Active\ Cases}{Exposed\ population}}{\frac{New\ Cases}{Exposed\ population}} \tag{2}$$

$$\text{Duration of disease} = \frac{\text{Active Cases of COVID} - 19}{\text{New Cases of COVID} - 19} \tag{3}$$

The models to estimate the Active Cases on i^{th} day was:

$$\widehat{Y} = Ae^{a_1 t + a_2 t^2} \text{ and } \widehat{Y} = Be^{b_1 t + b_2 t^2 + b_3 t^3} \tag{4}$$

where, $A = e^{a_0}$, $B = e^{b_0}$ and a_0, a_1, a_2, and b_0, b_1, b_2 and b_3 are the Regression Coefficients.

95% Confidence Interval (CI) of predicted active cases on each day was computed as:

$$95\%\ \text{CI} = e^{\text{Predicted Number oflog(Active Cases)} \pm 1.96 * \text{SE(Predicted number oflog(Active Cases))}} \tag{5}$$

$$95\%\ \text{CI} = e^{Ae^{a_1 t + a_2 t^2} \pm 1.96 * \text{sqrt}\left(\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}\right)} \tag{6}$$

where, S.E. for Quadratic model was computed as:

$$\text{S.E.(Predicted Number of Active Cases at } x_i \text{ day)} = \text{S.D.}\sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} + \frac{(x_i^2 - \bar{x}^2)^2}{\sum (x_i^2 - \bar{x}^2)^2}} \tag{7}$$

where S.D. is the Standard Deviation of y, which is calculated as:

$$\text{S.D.} = \sqrt{\frac{\sum (y - \hat{y})^2}{n - 1}} \tag{8}$$

where, y is the log transformation of seven days moving average of Total Active Cases, $\hat{y}$ is the predicted number of the Active Cases, and 'n' is the number of days Active Cases were observed. The coefficient of determination (R^2) is evaluated to assess the good fit of the model. The entire discussion and result section will be based on six regions in accordance with The National Family Health Survey (NFHS) division.

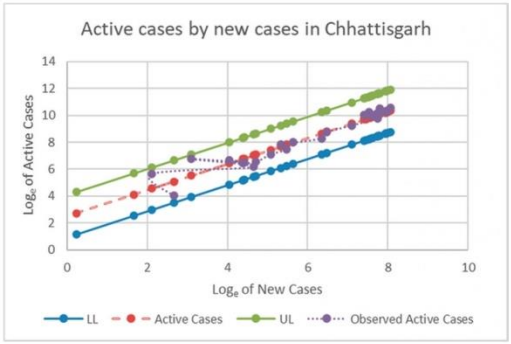


Fig. 1. Active cases by new cases in Chhattisgarh

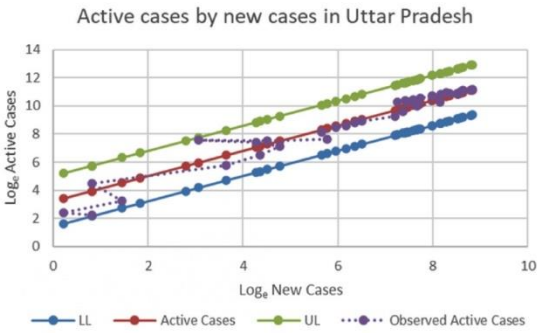


Fig. 2. Active cases by new cases in Uttar Pradesh

3. Results

The prediction model was performed with respect to each state. Histogram for the standardized residual of the fitted regression model with log of active cases as dependent variable looked plausible where residuals approximated a normal distribution with no weird shapes or low values or extremely high values. Normal P–P plot demonstrated that the distribution of standardized residual deviates only moderately from a distribution that was classically bell-shaped. The scatter plot indicated that the distribution of residuals across the expected values was equal,

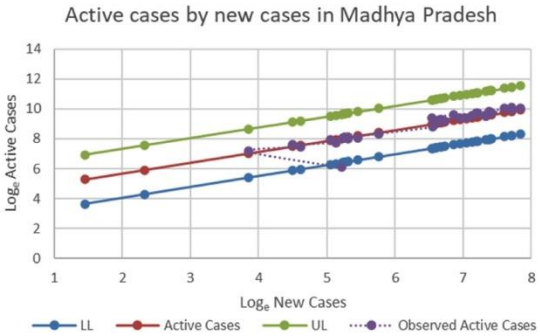


Fig. 3. Active cases by new cases in Madhya Pradesh

Table 1
Average recovery duration in days of COVID-19 patients in Indian states.

Region	State	R Square	F Statistic	B	b_1	Average Duration (Days)	Lower Limit	Upper limit
Central	**Chhattisgarh**	0.941	4184.161	2.511	0.969	12.319	10.406	14.583
	Uttar Pradesh	0.883	2101.031	3.222	0.896	25.073	19.353	32.484
	Madhya Pradesh	0.761	835.595	4.225	0.726	68.360	50.348	92.814
East	**Odisha**	0.956	5823.254	2.562	0.940	12.958	11.168	15.035
	West Bengal	0.948	4789.756	2.441	0.980	11.482	9.497	13.881
	Bihar	0.923	3143.014	2.808	0.914	16.569	13.637	20.132
	Jharkhand	0.871	1695.179	3.204	0.867	24.631	19.806	30.632
North	**Rajasthan**	0.937	4179.219	2.650	0.952	14.157	11.801	16.984
	Jammu and Kashmir	0.932	3737.557	2.604	1.007	13.518	11.337	16.118
	Himachal Pradesh	0.915	2802.949	2.914	0.898	18.424	15.930	21.309
	Uttarakhand	0.901	2441.060	2.986	0.899	19.815	16.625	23.616
	Haryana	0.891	2274.574	3.440	0.806	31.181	25.567	38.029
	Delhi	0.870	1884.836	3.123	0.892	22.704	17.157	30.043
	Punjab	0.802	1112.494	3.609	0.811	36.918	28.370	48.042
	Chandigarh	0.789	986.271	3.457	0.740	31.725	26.839	37.502
	Ladakh	0.750	829.768	3.337	0.806	28.140	23.760	33.327
Northeast	**Manipur**	0.936	3776.925	1.598	1.224	4.943	4.213	5.800
	Arunachal Pradesh	0.863	1568.943	1.838	1.193	6.283	5.073	7.783
	Meghalaya	0.856	1414.242	3.140	0.878	23.105	19.702	27.097
	Assam	0.850	1423.111	3.540	0.842	34.464	26.671	44.534
	Nagaland	0.671	417.574	2.869	0.980	17.625	12.427	24.999
	Sikkim	0.669	401.005	3.353	0.754	28.580	22.841	35.760
	Mizoram	0.588	368.827	2.743	0.955	15.540	12.208	19.782
South	**Tamil Nadu**	0.967	8117.149	1.911	1.061	6.758	5.710	7.999
	Karnataka	0.950	5227.651	2.804	0.968	16.514	13.753	19.829
	Kerala	0.946	5514.687	2.629	0.966	13.861	11.882	16.170
	Andhra Pradesh	0.898	2381.962	2.940	0.926	18.909	14.610	24.472
	Telangana	0.840	2767.206	2.936	0.935	18.836	15.647	22.675
	Puducherry	0.826	1257.392	3.133	0.878	22.945	18.743	28.089
	Andaman and Nicobar Islands	0.804	1052.745	1.967	1.139	7.150	5.993	8.530
West	**Gujarat**	0.920	3043.271	2.251	1.056	9.494	7.450	12.098
	Goa	0.884	1949.522	2.036	1.064	7.664	6.167	9.524
	Maharashtra	0.864	1746.580	3.102	0.965	22.240	15.389	32.142
	Dadra, Nagar Haveli, Daman and Diu	0.732	593.038	2.688	0.835	14.699	12.449	17.355

p value is < 0.05 for all states, B = unstandardized beta coefficient, b1 = slop of the regression line.

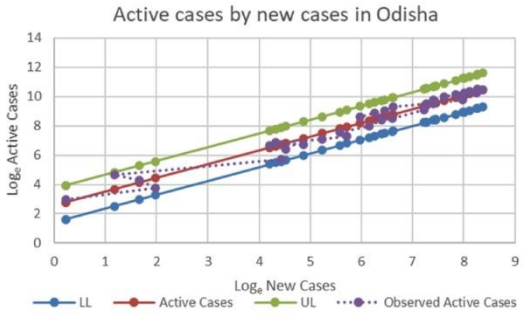


Fig. 4. Active cases by new cases in Odisha

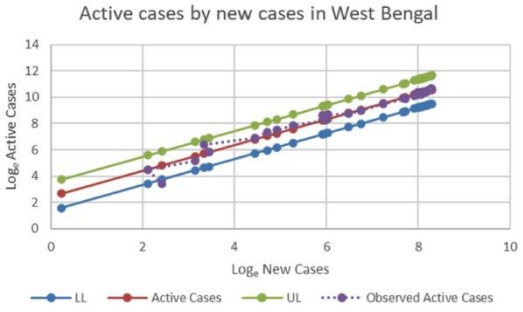


Fig. 5. Active cases by new cases in West Bengal

suggesting that the model was homoscedastic. Statistical models have been constructed for the duration of patients under COVID-19 across all states in India. The model fitted well for all the regions. Log of new cases and log of observed cases were plotted against the log of predicted active cases with the corresponding upper and lower confidence limits.

3.1. Central region

Log of new cases was plotted against log of predicted active cases with the corresponding upper and lower confidence limits and was found linear for central states of India comprising the states Chhattisgarh, Uttar Pradesh, and Madhya Pradesh (Figs. 1–3). Even though the model fit was enough to predict the average duration of COVID-19, there was a significant difference in the average duration between central states. Madhya Pradesh($R^2 = 0.76$) has an average duration as high as 68 days (CI = 50.348–92.814), whereas Uttar Pradesh($R^2 = 0.88$) with 25 days (CI = 19.35–32.48) and Chhattisgarh($R^2 = 0.94$) with 12 days (CI = 10.40–14.58) (Table 1).

3.2. Eastern region

Log of new cases against log of predicted active cases was plotted for Odisha, West Bengal, Bihar, and Jharkhand (Figs. 4–7). The average duration in days where a patient remains infected is identified as 11 days (CI = 9.49–13.88) in the state of West Bengal ($R^2 = 0.94$), while the average days increased to 25 days (CI = 19.80–30.63) in Jharkhand ($R^2 = 0.87$). Duration of disease in Odisha($R^2 = 0.95$) and Bihar($R^2 = 0.92$) are 13 days (CI = 11.16–15.03) and 17 days (CI = 13.63–20.13)

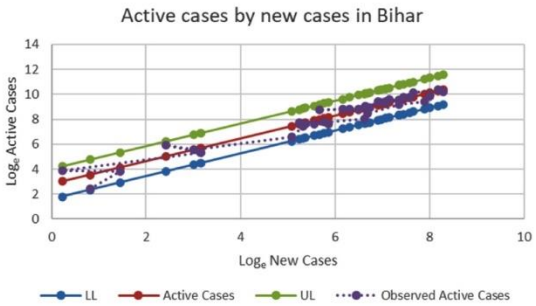


Fig. 6. Active cases by new cases in Bihar

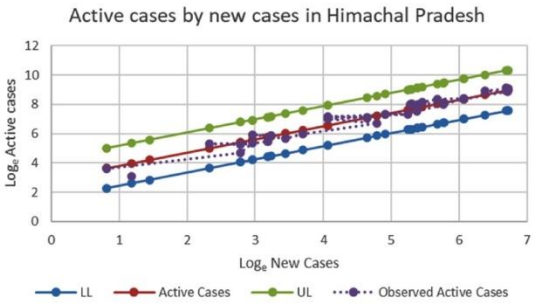


Fig. 10. Active cases by new cases in Himachal Pradesh

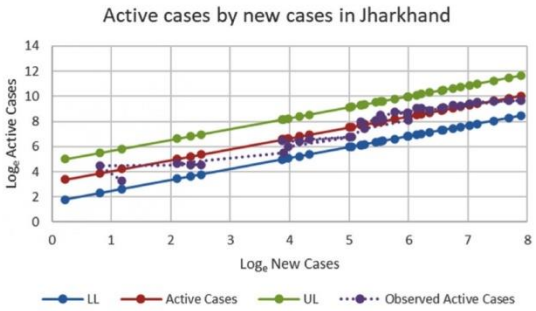


Fig. 7. Active cases by new cases in Jharkhand

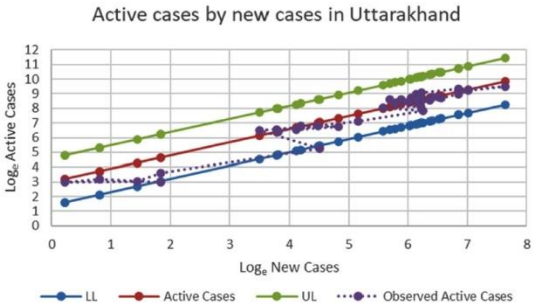


Fig. 11. Active cases by new cases in Uttarakhand

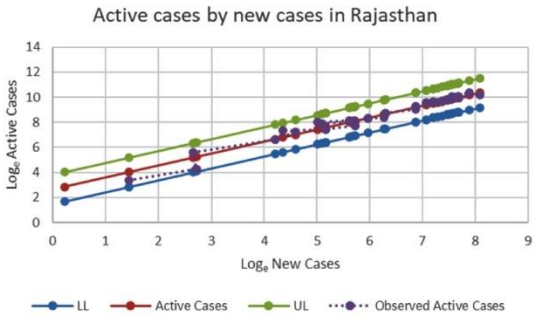


Fig. 8. Active cases by new cases in Rajasthan

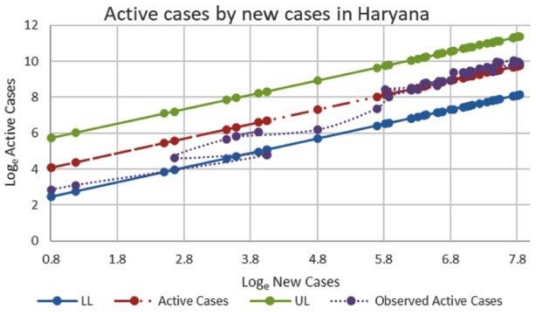


Fig. 12. Active cases by new cases in Haryana

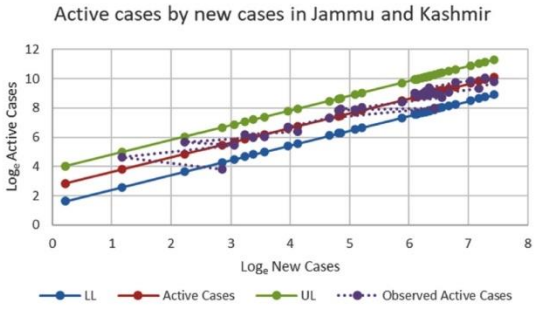


Fig. 9. Active cases by new cases in Jammu and Kasmir

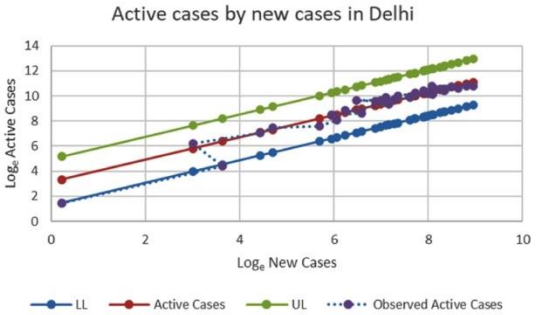


Fig. 13. Active cases by new cases in Delhi

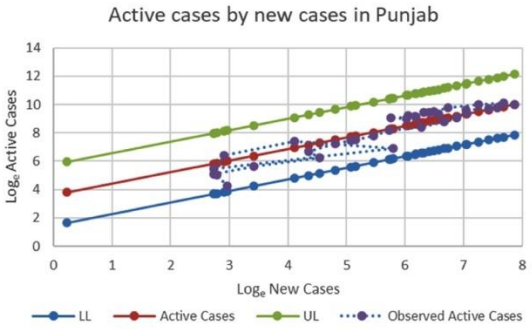


Fig. 14. Active cases by new cases in Punjab

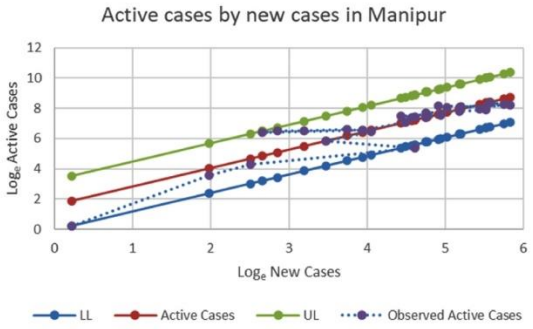


Fig. 17. Active cases by new cases in Manipur

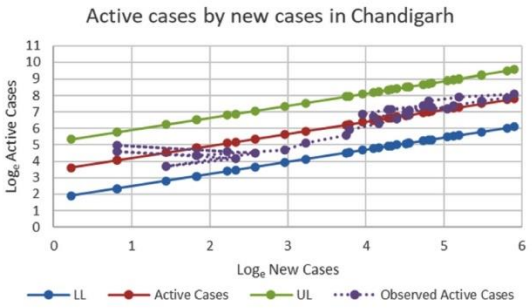


Fig. 15. Active cases by new cases in Chandigarh

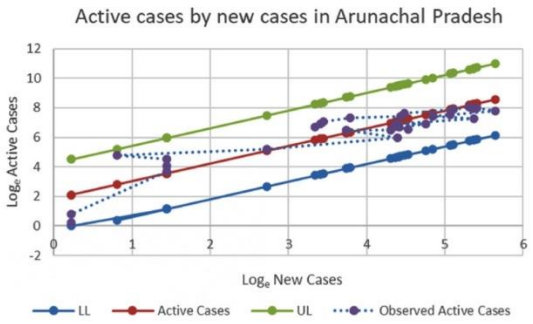


Fig. 18. Active cases by new cases in Arunachal Pradesh

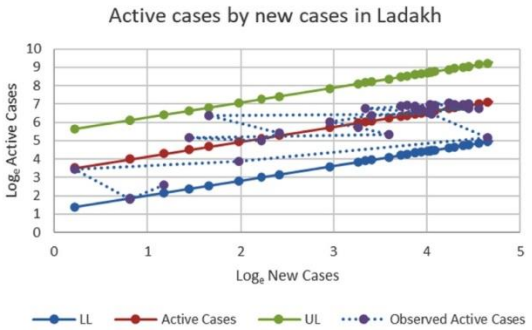


Fig. 16. Active cases by new cases in Ladakh

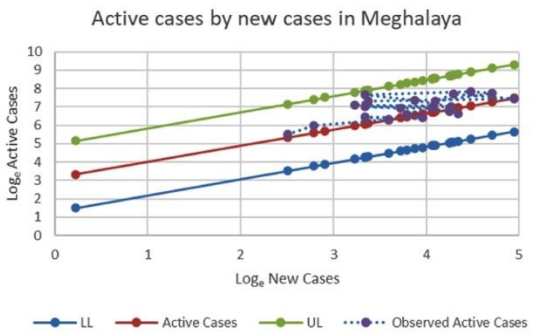


Fig. 19. Active cases by new cases in Meghalaya

respectively (Table 1).

3.3. Northern region

In the northern region, log of new cases against log of predicted active cases was plotted with the corresponding upper and lower confidence limits for the states Rajasthan, Jammu and Kashmir, Himachal Pradesh, Uttarakhand, Haryana, Delhi, Punjab, Chandigarh, and Ladakh and were found to be linear (Figs. 8–16). The average duration of days in Rajasthan ($R^2 = 0.93$) and Jammu Kashmir ($R^2 = 0.93$) is 14 days (CI = 11.80–16.98) (CI = 11.33–16.11), Himachal Pradesh ($R^2 = 0.91$) takes around 18 days (CI = 15.93–21.30) while Uttarakhand ($R^2 = 0.90$) takes around 20 days (CI = 116.62–23.61) followed by Haryana ($R^2 = 0.89$) 31 days (CI = 25.56–38.02) and Chandigarh ($R^2 = 0.78$) 32 days (CI = 26.83–37.50). With an average length of 36 days, Punjab ($R^2 = 0.8$) has the highest recovery rate among the northern states (CI = 28.37–48.04). Delhi ($R^2 = 0.87$) has an average recovery time of 22 days, having a confidence interval of 17 days–30 days. The union territories like Ladakh ($R^2 = 0.75$), which has a model fit of 75% ($R^2 = 0.75$) and has an average recovery duration of 28 days (CI = 23.76–33.32) (Table 1).

3.4. Northeastern region

Log of new cases against log of predicted active cases with the upper and lower confidence limits was plotted for northeast states, including Manipur, Arunachal Pradesh, Meghalaya, Assam, Nagaland, Sikkim, and Mizoram (Figs. 17–22). High model fit was found in states of Manipur

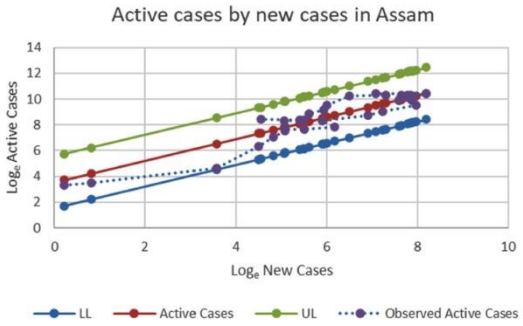


Fig. 20. Active cases by new cases in Assam

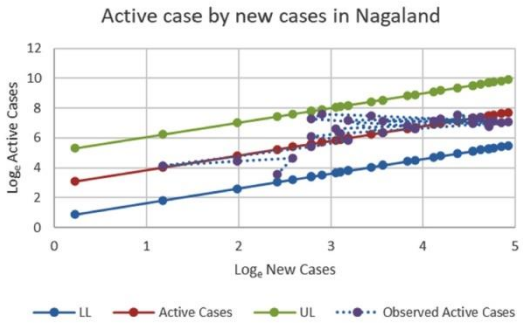


Fig. 21. Active cases by new cases in Nagaland

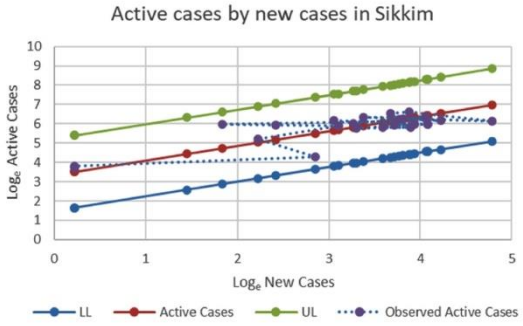


Fig. 22. Active cases by new cases in Sikkim

($R^2 = 0.93$), Arunachal Pradesh ($R^2 = 0.86$), Meghalaya ($R^2 = 0.85$) and Assam ($R^2 = 0.85$) with average recovery days of 5, 6, 23, 34 days (Table 1). Nagaland ($R^2 = 0.67$) has a model fit of 67% ($R^2 = 0.67$) and has a duration of 18 days (CI = 12.4–24.9), similarly Sikkim($R^2 = 0.66$) with a model fit of 67% ($R^2 = 0.67$) and duration of 29 days (CI = 22.84–35.76). Finally, Mizoram ($R^2 = 0.58$) has the least model fit with 59% ($R^2 = 0.59$) and a duration of 16 days (CI = 12.20–19.78) (Table 1).

3.5. Southern region

For southern states named Tamil Nadu, Karnataka, Kerala, Andhra Pradesh, with the union territories of Puducherry, and Andaman and

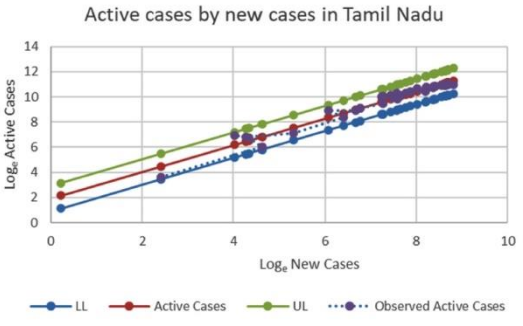


Fig. 23. Active cases by new cases in Tamil Nadu

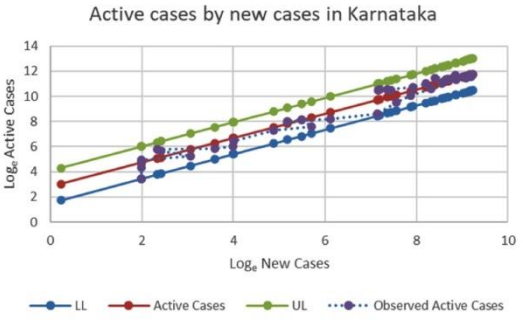


Fig. 24. Active cases by new cases in Karnataka

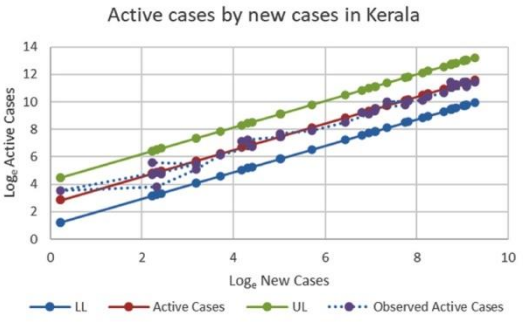


Fig. 25. Active cases by new cases in Kerala

Nicobar Islands, log of new cases against log of predicted active cases with the upper and lower confidence limits were plotted (Figs. 23–28). Andhra Pradesh ($R^2 = 0.89$) takes an average of 19 days (CI = 14.61–24.47, 15.64–22.67) to recover, where Andaman and Nicobar Islands ($R^2 = 0.8$) and Tamil Nadu ($R^2 = 0.96$) take an average of 7 days (CI = 5.99–8.53, 5.71–7.99) to recover, followed by Kerala ($R^2 = 0.94$) with 14 days (CI = 11.88–16.17), Karnataka ($R^2 = 0.95$) with 17 days (CI = 13.75–19.82), and Puducherry ($R^2 = 0.82$) with 23 days (CI = 18.74–28.08) (Table 1).

3.6. Western region

In western states, Gujarat, Goa, Maharashtra Dadra, Nagar Haveli, Daman and Diu, log new cases against log predicted active cases along

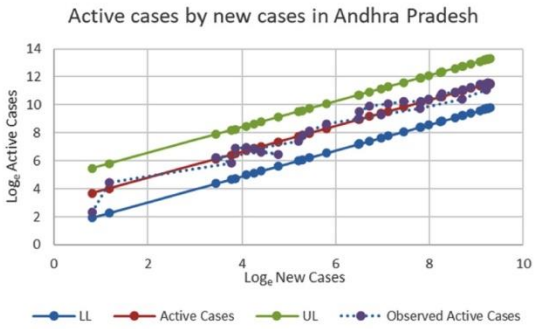


Fig. 26. Active cases by new cases in Andhra Pradesh

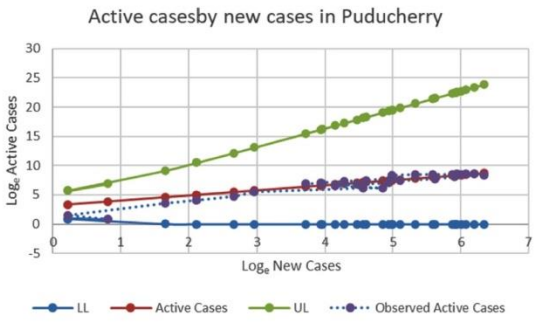


Fig. 27. Active cases by new cases in Puducherry

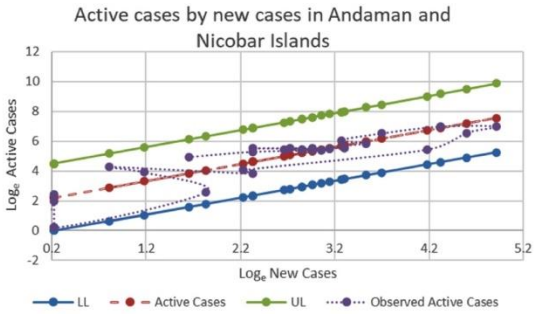


Fig. 28. Active cases by new cases in Andaman and Nicobar Islands

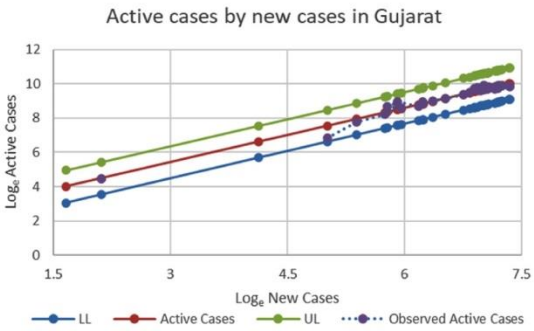


Fig. 29. Active cases by new cases in Gujarat

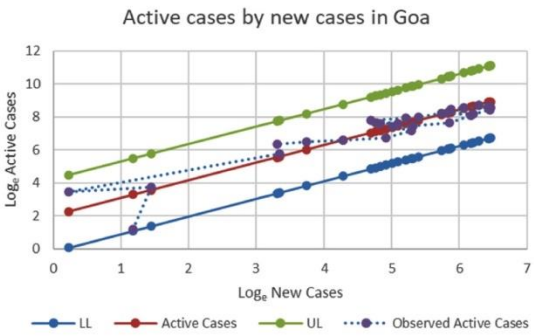


Fig. 30. Active cases by new cases in Goa

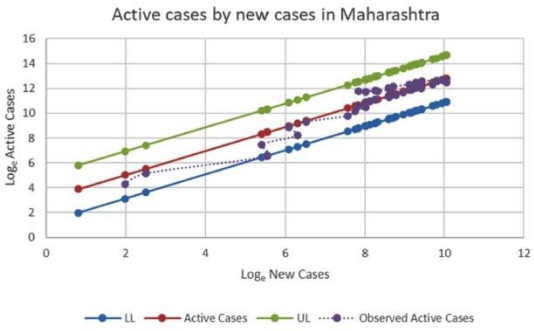


Fig. 31. Active cases by new cases in Maharashtra

with its upper and lower confidence limits was plotted (Figs. 29–32). Dadra and Nagar Haveli and Daman and Diu ($R^2 = 0.73$) have a model fit of 73% ($R^2 = 0.73$) and have a duration of 15 days with a confidence interval of 12.44 days and 17.35 days wherein Goa ($R^2 = 0.88$) and Gujarat ($R^2 = 0.92$) the duration is almost equal, almost 8–10 days (CI = 7.45–12.09, 6.16–9.52). Maharashtra ($R^2 = 0.86$) has an average recovery time of 22 days, with a confidence interval of 15 days–32 days (Table 1).

4. Discussion

The model suited ideal for most states, some union territories and northeastern states considered to have a reduced fit in identifying the average time. The average time of recovery from the disease in India ranges from 5 days to 68 days. However, Hospital studies by Manash et al. (2020) showed that COVID-19 patients in India have an estimated recovery time of 25 days (CI = 16 days–34 days). The recovery period estimated for male and female patients and patients belonging to different age groups is also not statistically significant.[9] The rigorous proof for the relation between age-independent incidence, prevalence and time duration was initially provided by Keiding (1991). Miettinen (1976), Freeman and Hutchison (1980) had earlier presented heuristic derivations for the same special case. A related identity has also been proven by Keiding (1991) (see Freeman and Hutchison, 1980, p. 709). The relationships between time period and cohort measures under general growth were addressed by Preston (1987).[3] The relationship between prevalence, incidence, and disease duration is studied in a

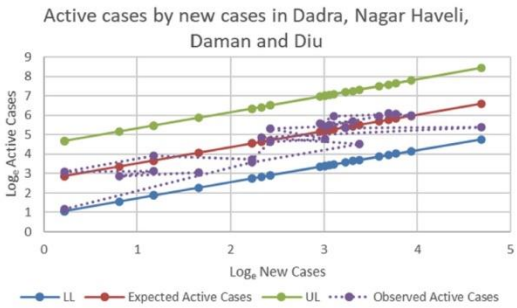


Fig. 32. Active cases by new cases in Dandra, Nagar Haveli, Daman and Diu

linear relationship for a stable population. The nationwide lockdown due to COVID-19 has produced the closed population's situation where migration is considered negligible.[4] The average recovery time in Madhya Pradesh (approx. 68 days) and Manipur (approx. 5 days) was found to be outliers when compared with other states in India. As per Lancet, Madhya Pradesh and Telangana found a high overall vulnerability (index value more than 0.75) for the management of and response to the COVID-19.[8]

In India, the first positive case was registered in Kerala on January 27, 2020; however, the first case of COVID-19 in North East India was found precisely in Manipur, which was a bit late compared to other regions of India.[15] Studies showed that Assam and Meghalaya had a very low potential compared to Tripura in controlling the present pandemic situation. A moderate potentiality to control the situation was found in Arunachal Pradesh and Nagaland. Sikkim and Mizoram have a very high potentiality than Manipur to control COVID-19.[16] In this study, the model fits best for the states of Manipur, Arunachal Pradesh, Meghalaya and Assam, with average recovery days of 5, 6, 23, 34 days correspondingly and reduced fit for Nagaland with 18 days, Sikkim with 29 days. Mizoram has the least fit and an average duration of 16 days.

On the west side of India, Dadra and Nagar Haveli and Daman and Diu have the least model fit and have a duration of 15 days, whereas Gujarat, Goa, and Maharashtra have a high model fit the duration is almost equal, almost 9, 7, and 22 days correspondingly. Testing in the Union territories became functional only on April 6, 2020, after which 23,023 samples have been tested to date January 21, 2021. Center explicitly enforced lockout, door-to-door survey and screening were scheduled, and 400 teams along with personnel from the Health and Education Department were dispatched to the field to perform comprehensive health screenings.[6]

The significant finding from the study by Kapil Ghosh et al.(2020) shows the inaccessibility of testing centres in the Eastern (Odisha, West Bengal, Bihar, and Jharkhand) Central and extreme Northern portion of India.[17] The average duration of days where a patient remains in COVID-19 is approximately 11, 13, 17, 25 days in West Bengal, Odisha, Bihar, and Jharkhand. Every state in the eastern region has a good model fit. While central states like Chhattisgarh, Uttar Pradesh has a moderate fit with 12 and 25 average recovery days. Madhya Pradesh is considered as an outlier with a reduced fit and average recovery days of 68. Assessing the northern region, the average duration of days in Rajasthan and Jammu Kashmir is 14 days; Himachal Pradesh takes around 18 days while Uttarakhand takes around 20 days, followed by Haryana 31 days and Chandigarh 32 days. The maximum duration taken by a state in the northern region is Punjab, with 36 days. The union territories like Ladakh, which has a reduced model fit and has an average duration of 28 days for a patient to be in the disease. Goel K et al. (2020) has studied and found that Srinagar-Leh highway opening and Iran pilgrimage returnees were the major challenges to estimate the actual data of COVID patients. The study also suggested that there is a need for a more robust surveillance system.[5] Lockdown as restricting the migration of population and decision taken by the government to quarantine the population may greatly reduce the risk of continued spread of the epidemic in India.[18] Towards the south, Andhra Pradesh takes approximately an average of 19 days to recover. Andaman and Nicobar Islands and Tamil Nadu takes an average of 7 days correspondingly for the recovery, followed by Kerala with 14 days, Karnataka with 17 days, and Puducherry 23 days.

5. Conclusion

Tamil Nadu was identified as the best fit for the model with an R^2 value of 0.96, and Mizoram was identified to have the least fit with an R^2 value of 0.58. At the same time, Manipur was identified to have the least average recovery duration with 4.94 days and the highest in Madhya Pradesh with 68 days. Despite the fact that the data was derived from government sources, the data quality and patient follow-up were not ensured by the database. Government and the hospitals' policies in admitting, fear in the community of getting admitted to a hospital and the healthcare workers' behaviour towards the COVID-19 patients, sensitivity and specificity of the diagnostic tests may also be some factors that could have caused the outlying in cases of Madhya Pradesh and Manipur. To an extent, the patients remaining in-home and migration in each state's population is identified as a limitation to the study. However, most of the states have different situations concerning COVID-19 and have unique problems to tackle, so it is preferred to consider these situations while comparing each state's average duration.

Conflict of interest

We hereby declare that the article entitled "**COVID-19 pandemic and its average recovery time in Indian states**" does not have any conflict of interest involved.

We also declared that this research article is nowhere submitted for publication.

Ethical approval and consent to participate

Not applicable.

Availability of supporting data

Open repository of Github.

Funding

No funding received.

Declaration of competing interest

The authors declare that they have no conflict of interest.

References

1 Rees EM, et al. COVID-19 length of hospital stay: a systematic review and data analysis, 18 https://doi.org/10.1186/s12916-020-01726-3; 2020, 270.
2 Lauer SA, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Intern Med.* 2020; 172(9):577–582. https://doi.org/10.7326/M20-0504.
3 Alho JM. On prevalence, incidence, and duration in general stable population. *Biometrika.* 1992;48(2):587–592. https://doi.org/10.2307/2532312.
4 Kumar R, Mishra RS. Estimation of COVID – 19 prevalence under closed population using capture – recapture approach. *Int J Sci Res.* 2020;9(7):451–457. https://doi.org/10.21275/SR20629043404, 2020.
5 Goel K, et al. The successful containment of COVID-19 outbreak in Union Territory of Ladakh, India. *J Fam Med Prim Care.* 2021;9(11):5574–5579. https://doi.org/10.4103/jfmpc.jfmpc_1413_20.
6 http://www.businessworld.in/article/COVID-19-Daman-Diu-And-Dadra-Nagar-Haveli-Script-A-Mini-Success-Story/16-06-2020-287875/.

7 Khan Q, Swaminathan J. Impact of COVID-19 lockdown on mental health and social life of university students of Delhi. *Int J Health Sci Res*. 2020;10(12):222–230.
8 Acharya R, Porwal A. A vulnerability index for the management of and response to the COVID-19 epidemic in India: an ecology study. *Lancet*. 2020;8(9):1142–1151. https://doi.org/10.1016/S2214-109X(20)30300-4.
9 Barman MP, Rahman T, Bora K, Borgohain C. COVID-19 pandemic and its y time of patients in India: a pilot study, Diabetes & Metabolic Syndrome. *Clin Res Rev*. 2020. https://doi.org/10.1016/j.dsx.2020.07.004.
10 Keiding N. Age-specific incidence and prevalence: a statistical perspective. *J Roy Stat Soc*. 1991;Series A 154:371–412.
11 Miettinen OS. Estimability and estimation in case-referent studies. *Am J Epidemiol*. 1976;103:226–235.
12 Freeman J, Hutchison GB. Prevalence, incidence, and duration. *Am J Epidemiol*. 1980;112:707–723.
13 Preston SH. Relations among standard epidemiologic measures in a population. *Am J Epidemiol*. 1987;126:336–345.
14 Chatterjee Patralekha. Is India missing COVID-19 deaths? *Lancet*. 2020;396(10252). https://doi.org/10.1016/S0140-6736(20)31857-2. Page 657, ISSN 0140-6736.
15 Ghosh P, Ghosh R, Chakraborty B. COVID-19 in India: statewise analysis and prediction. *JMIR public health and surveillance*. 2020;6(3), e20341. https://doi.org/10.2196/20341.
16 Mahato R, Bushi D, Nimasow G. AHP and GIS-based risk zonation of COVID-19 in North East India. *Curr World Environ*. 2020;15(3). https://doi.org/10.12944/CWE.15.3.29.
17 Ghosh Kapil, Sengupta Nairita, Manna Dipanwita, Sunil Kumar De. Inter-state transmission potential and vulnerability of COVID-19 in India. *Progress in Disaster Science*. 2020;7. https://doi.org/10.1016/j.pdisas.2020.100114, 100114,ISSN 2590-0617.
18 Bhattacharjee Atanu, Kumar Mukesh, Patel Kamalesh Kumar. When COVID-19 will decline in India? Prediction by combination of recovery and case load rate. *Clinical Epidemiology and Global Health*. 2021;9:17–20. https://doi.org/10.1016/j.cegh.2020.06.004. ISSN 2213-3984.

Statistical Model for COVID-19 in Different Waves of South Indian States

Dialogues in Health 1 (2022) 100016

Contents lists available at ScienceDirect

Dialogues in Health

journal homepage: www.elsevier.com/locate/dialog

Statistical Model for COVID-19 in Different Waves of South Indian States

Noel George *, Jang Bahadur Prasad, Pradyuman Verma

Department of Epidemiology and Biostatistics, KLE University, Belgaum- 590010, Karnataka, India,

ARTICLE INFO

Keyword:
COVID-19
Active Cases
Case Fatality Rate
Phases
South India

ABSTRACT

Background: COVID-19 has resurfaced in India, where it is rapidly spreading and wreaking havoc in rural areas. An effort has been undertaken to assess the levels and patterns of COVID-19 active cases in the southern states of India. To trace and reason out anomalous trends in the COVID-19 curve so that particular actions such as lockdown, de-lockdown, and healthcare improvisation can be implemented at the appropriate time.

Methods: The data has retrieved from the government websites through a platform called Kaggle. The entire duration of COVID – 19 were classified into three compartments: Phase one, Resting phase, and Phase two. The Case Fatality Rate in south Indian states was analysed corresponding to the phases, and a compartmental model for COVID-19 dynamics in the region was proposed.

Results: The quadratic regression model was fitted and found to be the best model for the phases except for the resting phase. Phase one was comparatively less fitted when compared to phase two. In most of the south Indian states, the active cases in phase one were almost more than four times that of phase two. The average CFR value in phase one was lower than the subsequent phase in all of the southern Indian states. In phase one, Telangana, Karnataka, and Tamil Nadu had the highest CFR (4.77,4.22, and 3.71, respectively), whereas Lakshadweep and Kerala had the lowest CFR (0.27 and 0.71, respectively). In the resting phase, the CFR stabilized in all states and reached a value between 0.2 to 2. The trend was similar in phase two also, CFR of Lakshadweep, Kerala, Telangana, and Andhra Pradesh (0.143, 0.416,0.553, 0.803) were very low, while the CFR of Andaman and Nicobar Islands, Karnataka, and Tamil Nadu (1.237, 1.306, 1.490) were very high.

Conclusion: The first and second phases of the COVID-19 virus in south Indian states had different characteristics. A District-level working group with the autonomy to respond to rapidly changing local situations must be empowered to tackle the next phase. The upcoming phases could be more peaked in less time and could be a hectic situation for the health care system.

Introduction

COVID-19, which began as a pneumonia-like outburst in late 2019 in the city of Wuhan, has spread globally to become a catastrophic event.[1] The first instance of COVID-19 in India, specifically in the southern states of Kerala, was reported on January 30, 2020[2], and it has since impacted over thirty million individuals in the country, resulting in about four hundred thousand deaths. The physical manifestation comprises respiratory illness with symptoms that range from a common cold to severe viral pneumonia, which causes acute respiratory distress, typically fatal in nature. The first wave peaked in September 2020, with daily cases reaching about 0.1 million. The morbidity was initially low, but it becomes noticeable towards the middle of June 2020. The decrease in the first wave was most likely due to several factors, including effective government actions, increased awareness, and, most crucially, medical professionals' expertise in treating the disease in the early months.

The second COVID-19 wave in India, which was different in different states roughly started on February 11, 2021, has been disastrous, with a number of cases reaching 0.2 million per day on April 14, 2021, which is already quadruple that of the first wave peak. The first and second waves are around five months apart. The rapid increase in the number of cases after such a long period of 'cooling' is baffling and wreaking havoc across the globe. This wave also spread to rural communities and villages of India as of May 2021.[3] States like Delhi, Karnataka, Tamil Nadu, Kerala, and Maharashtra are severely affected by the second wave.[4] The number of cases and deaths have risen drastically in India. Based on the cases discovered by the USA and Brazil on April 10, 2021, India has been in the third position.[5] According to an exponential fit on the current data, the infection rate is substantially more significant than the previous wave, but the case fatality rate is lower. Preliminary projections using the SIR model estimates that the second wave will peak in mid-May 2021, with daily counts topping 0.35 million.[6]

* Corresponding author.
E-mail address: noel@pvalue.co.in (N. George).

http://dx.doi.org/10.1016/j.dialog.2022.100016
Received 5 August 2021; Received in revised form 22 May 2022; Accepted 22 May 2022
Available online 27 May 2022

The viral genome has acquired mutations over time as it goes throughout its host and results in the infamous double mutant COVID variant. India has almost 7,684 variations of COVID-19 virus.[7] The current increase in the number of infections in some Indian states may also be due to virus variants. South African, The UK, Brazilian, and Nigerian variants and variants that bear the lineage of such variants are among the critical spike mutation of worldwide concern that is a key for surveillance in the Indian landscape. N440K, a mutation with higher frequencies in Andhra Pradesh, is also included in the list. Despite the increased caseload, numerous national movements, such as the farmers' movement, have been active, and elections have been held in some states. These factors increase the risk of COVID-19 transmission.[8]

During the first wave, the Indian government imposed nationwide curfews, social distancing, and lockdown, which further implemented as a strict lockdown and other countermeasures so that infected individuals are identified and isolated.[9] However, because the second wave differed from the first, the government decided to speed up the vaccine campaign to combat COVID-19 with a state-wise strict lockdown. The government takes urgent steps along two paths of action: large-scale vaccination to reduce disease severity and break the transmission chain through safe behaviours.

The overall Case Fatality Ratio (CFR) has been estimated to be 1.3 percent since the pandemic began in March 2020, but the CFR among individuals who have infected the virus since 2021 is much lower at 0.87 percent. Provisionally, CFR looks to be lower in the second wave. Despite this, India reports 664 deaths a day (7-day moving average as of April 10, 2021). Fatality number lags the infection rate, and it is likely to increase as infectious rise.[10] The average time to recover from disease was ranging from 5 to 36 days in Indian states/union territories.[11]

Studies assessed lockdown extension by fitting an exponential model to the rise of COVID-19 using data from multiple states (Mondal and Ghosh, 2020).[12] Since no infection/ disease could ever follow the exponential distribution due to the virus's self-diminishing properties, modelling the growth and forecasting the number of affected persons and deaths is difficult; though there is a long tradition of research in the epidemiology of diseases in statistical modelling (Pastor-Satorras et al., 2015).[13] To ease the strain of Healthcare System and provide the best possible care for COVID-19 patients, appropriate models are expected to estimate the risk of Active Cases for proper planning, diagnosis, and cure of the disease.[14] Therefore, it is imperative to study the many phases of each wave, as the consequences of attempting to manage COVID-19, on the other hand, are being felt not only in the health sector, but also in the environment, the economy, and society. Every disease, more so communicable acute disease, is expected to follow a pattern, which usually includes, i) Acceleratory phase, ii) Optimum level, and ii) De-acceleratory phase. Usually, when data is available at the time of evaluation, the disease is expected to be in de-acceleratory phase. India's calculated case fatality rate has decreased to 1.59 percent as of September 21, 2020, from 3.38 percent in April.[15] From a public health point of view, it is critical to track case fatality rates and new cases over time in order to assess illness severity and reduce the degree of risk in disease pandemics. This study's motivation stems from a desire to add to the statistical investigation of COVID-19 pattern in South Indian states, where there is a paucity of literature. Hence, an attempt has been undertaken to study the levels and trends of the COVID-19 active cases, new cases, and CFR in southern states of India and to fit a quadratic regression model predict the upcoming phases of the pandemic so that appropriate actions as lockdown, de-lockdown and standardization of line of treatment could be taken at the appropriate time to optimize health and financial benefits.[16]

Materials and Methods

The COVID-19 data have been retrieved from the government website (https://www.mohfw.gov.in) through Kaggle (COVID-19 in India, 2020)[17], which provides a platform for extracting data on coronavirus infections to assist the worldwide community and health organizations in making better decisions. The website has hosted a number of tasks aimed at improving knowledge on COVID-19 infectious disease.

The state-wise trends of COVID-19Active Cases have been studied in South Indian states.

$$Active\ cases = Total\ Cases - (Cured + Death)$$

Case Fatality Rate of COVID-19 given as

$$\mathrm{CFR} = \frac{\mathrm{Deaths}}{\mathrm{Cured} + \mathrm{Deaths}} \times 100$$

The data on the number of confirmed cases, cured, and deaths available in the public domain from January 30, 2020 to May 19, 2021, were analysed using statistical software like SPSS-22 and Microsoft Excel 2016. The model was built using data from the day on which there were more than one death due to COVID-19. Since the active cases for some states/ union territory are not linearly related, transformations are used to make it easier to model and analysing data.

A quadratic regression model is a polynomial of degree 2 with all the terms present.

$$EY_t = \beta_0 + \beta_1 t + \beta_2 t^2$$

This equation has the linear effect parameter β1 and quadratic effect parameter β2, respectively as well as the constant parameter β0.[18]

The paper proposes a compartmental model for the dynamics of COVID-19 in south India. First of all, the data was normalised to the total population of individuals, and the compartments were created (Phase one, Resting Phase and Phase two) as per trends in Active Cases. The entire period of COVID in south Indian states has been divided into three groups. More specifically, the compartment collects all the individuals that are affected by the virus. Phase one is defined with the trend in active cases from the initial period, which has at least one extreme peak to the lowest value of the active cases. The resting phase is considered as the linear, less drastic period with an approximately constant magnitude in active cases. Phase two initiate from the sudden increase in the active cases from the restating phase.

Results

Table 1 reveals treatment performance of COVID-19 cases by states/ union territories. It shows that Karnataka and Lakshadweep have 25 percent active cases followed by Puducherry (19.92%), Kerala, Andhra Pradesh, and Tamil Nadu by 15%. In contrast, Telangana and Andaman and Nicobar Islands have only around 9 and 3 percentage of active cases respectively and could say the best in controlling the pandemic. Furthermore, the trends in the South Indian States' performance in controlling the COVID-19 have been analysed using a statistical model so that unusual trends in the cure of COVID-19 are traced and reasoned out.

Table 2 classifies the duration of COVID-19 disease in each state and has been classified into three phases. In each phase, there was a difference in the average Case Fatality Rate. The average Case Fatality Rate in most of the states are less than two. However, in phase one there was a change in

Table 1
COVID-19 Cured, Deaths and Confirmed cases and active cases in south Indian states of India, as of May 19, 2021

States	Cured	Deaths	Confirmed	Active Cases (%)
Andaman and Nicobar Islands	6359	92	6674	223 (3.34)
Andhra Pradesh	1254291	9580	1475372	211501 (14.34)
Karnataka	1674487	22838	2272374	575049 (25.31)
Kerala	1846105	6612	2200706	347989 (15.81)
Lakshadweep	3915	15	5212	1282 (24.6)
Puducherry	69060	1212	87749	17477 (19.92)
Tamil Nadu	1403052	18369	1664350	242929 (14.6)
Telangana	485644	3012	536766	48110 (8.96)

Table 2
Case Fatality Rate in south Indian states of India, as of May 19, 2021

State	Phase	Days	Minimum	Maximum	Mean	Std. Deviation
Andaman and Nicobar Islands	Phase1	183	0.508	4.313	1.568	0.578
	Resting phase	57	1.231	1.244	1.237	0.004
	Phase2	51	1.162	1.426	1.229	0.058
Andhra Pradesh	Phase1	178	0.813	2.879	1.493	0.721
	Resting phase	112	0.806	0.813	0.808	0.002
	Phase2	58	0.758	0.818	0.803	0.014
Karnataka	Phase1	307	1.310	37.500	4.228	5.364
	Resting phase	33	1.304	1.310	1.306	0.002
	Phase2	76	1.292	1.394	1.320	0.031
Kerala	Phase1	406	0.000	8.000	0.715	0.897
	Resting phase	13	0.414	0.419	0.416	0.002
	Phase2	52	0.357	0.432	0.414	0.020
Lakshadweep	Phase1	30	0.163	0.408	0.272	0.078
	Resting phase	18	0.134	0.160	0.143	0.007
	Phase2	35	0.075	0.394	0.238	0.114
Puducherry	Phase1	165	1.685	6.034	2.596	0.944
	Resting phase	112	1.664	1.691	1.676	0.007
	Phase2	65	1.616	1.730	1.668	0.028
Tamil Nadu	Phase1	265	0.000	50.000	3.713	7.174
	Resting phase	111	1.467	1.525	1.490	0.015
	Phase2	63	1.263	1.467	1.393	0.076
Telangana	Phase1	259	0.000	87.500	4.779	11.515
	Resting phase	97	0.547	0.578	0.553	0.007
	Phase2	68	0.549	0.634	0.568	0.021

the average Case Fatality Rate in almost all states thereafter the resting phase, and phase two had CFR less than 2.

Table 3, Quadratic Regression Models for COVID-19 by the States of India, reveals explanatory power of the Regression Models varying from 55% in phase 1 to 99% in phase two of Puducherry. The entire period of COVID 19 have been classified into three Phases, and their quadratic regression models were fitted. In the first Phase, the best-fitted state was Karnataka, followed by Andaman and Nicobar Islands. Whereas in the second phase, the best-fitted state/ union territory was Puducherry, followed by Tamil Nadu and Andaman and the Nicobar Islands. All the states had a linear trend with a slight slope towards the x-axis in the resting phase, and the R square values were found to be more than 90% in all the states.

In the first phase, the case fatality rate in every South Indian state/union tertiary fluctuated, then stabilized in the second and third phases. The CFR increased and gained prominence in the early stages of phase 1 before stabilizing in the range of 0.5 to 2. In Figure 1a, the patterns stabilized around the 160th day on the initial occurrence of the disease in Andaman and Nicobar Islands whereas in Andra Pradesh, it was around 180 days as per Figure 2a. In Karnataka and Kerala, CFR stabilized in 70 to 100 days, with the exception of a few heaps and dips as indicated in Figure 3a and Figure 4a . Figure 5a shows that the CFR rates in the union territory of Lakshadweep were relatively less as compare to other states or union territories, but to stabilize to a linear pattern was relatively less. In the phase 1 and resting phase, it showed a decreasing trend hitting below 0.1, and thereafter it increased and attained up to 0.4. Figure 6a of Puducherry shows a stability in CFR around 1.8 by the end of the first phase and continued the pattern throughout the resting phase and phase 2. Figure 7a depicts thatTamil Nadu was having a drastic increase in CFR in the initial stages of Phase 1, and thereafter by the 66th day, it also attained stability in the upcoming phases. In Figure 8a, Telangana also showed a similar trend to that of Karnataka and Kerala, which stabilized by the 75th day with minor fluctuations.

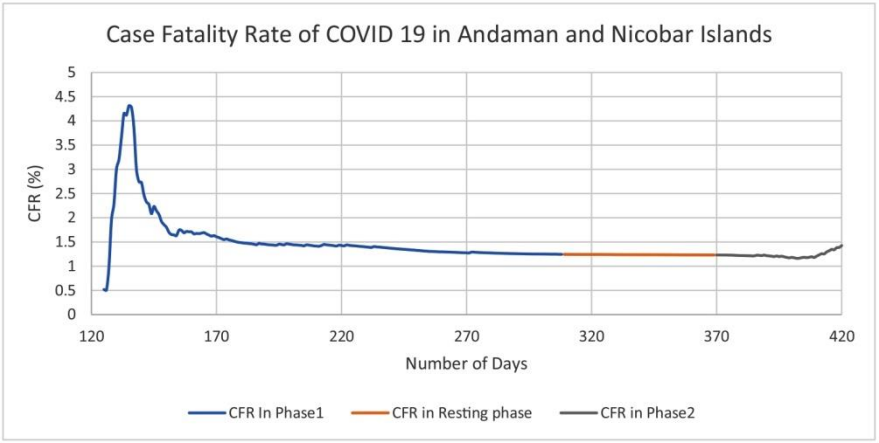
New cases and active cases had a nearly identical trend in every state, but the smoothness in active cases was much higher than in new cases. The first significant increase in the Andaman and Nicobar Islands occurred around the 110th day, but there was a minor trigger phase between 90 and 110 days. The first phase peaked around the 150th day (approximately 1180 Active Cases and 150 New Cases and abruptly declined during the next 25 days. The final quarter of phase one was slow and protracted, with a lot of fluctuations. The active and new cases stabilized with a small magnitude during the resting phase. Phase two began around the 309th day and increased until the 400th day, but it did not reach the same heights (approximately 210 Active Cases and 100 New Cases) as the previous phase (Figures 1b, 1c).

The peak of the Second phase (approximately 0.22 million Active Cases and 25thousand New Cases) in Andra Pradesh has clearly doubled that of Phase one(approximately 0.1 million Active Cases and 12 thousand New Cases). Phase one triggered around the 100th day and reached its maximum height by 180th day, and by 265, it was found to be slowly changing into a resting phase. The resting phase continued around 100 days, and thereafter a sudden trigger resulted in phase two by the 377th day of the initial COVID reporting (Figures 2b, 2c).

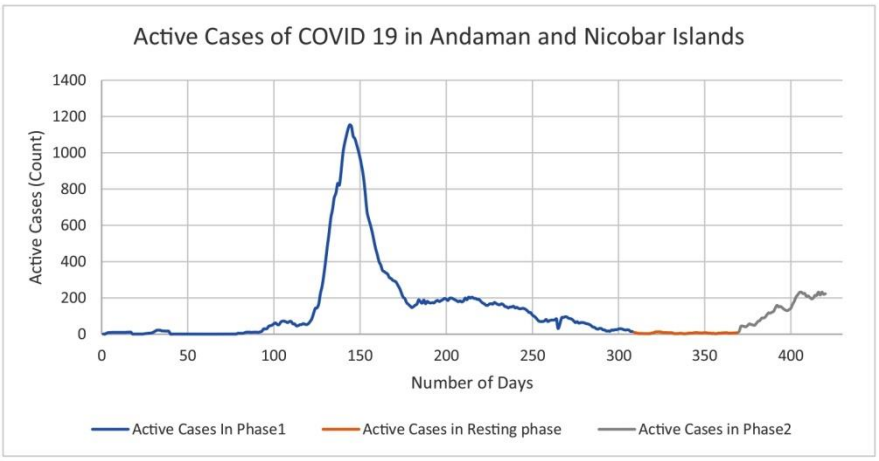
Karnataka had a drastic increase in the active and new cases when comparing phase one and two, it was around fivefold of the first phase. The resting phase found to be very less, around 33 days when compared to other states except for Kerala. Phase one increased drastically after the 100th day, and went up to the 250th day, then gradually decreased and went to a small resting period by the 328th day of initial infection. By 361thday the cases had increased and reached a maximum by 425th day (approximately 0.6 million Active Cases and 50 thousand New Cases)

Table 3
Quadratic Regression Models for COVID-19 instates of South India, as of May 19, 2021

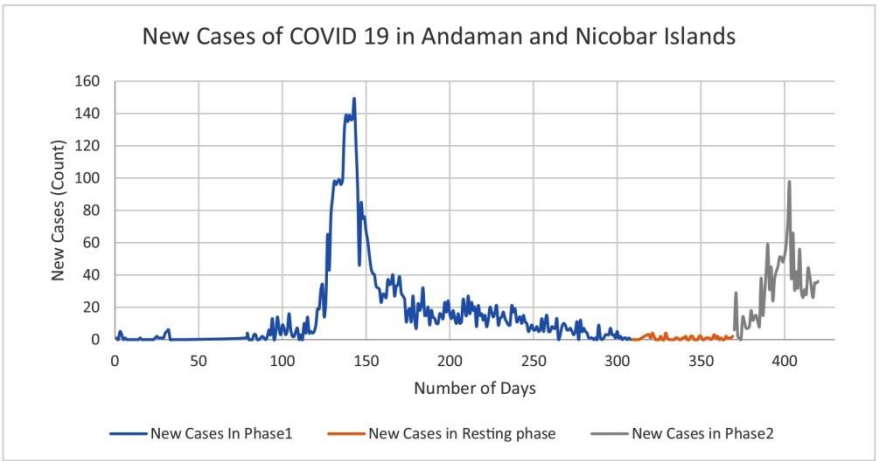
State	Phase	Model Summary					Parameter Estimates		
		R Square	F	df1	df2	Sig.	Constant	b1	b2
Andaman and Nicobar Islands	Phase 1	0.787	345.620	2	187	.000	0.089	-0.001	2.688E-06
	Phase 2	0.939	370.808	2	48	.000	-7746.987	35.594	-0.040
Andhra Pradesh	Phase 1	0.777	304.471	2	175	.000	-321699.471	4476.447	-12.535
	Phase 2	0.976	1121.953	2	55	.000	10397781.149	-55346.070	73.627
Karnataka	Phase 1	0.968	4906.606	2	324	.000	1.544	0.095	0.000
	Phase 2	0.975	1404.113	2	73	.000	27591727.516	-146626.706	194.737
Kerala	Phase 1	0.670	409.900	2	403	.000	-19787.861	293.836	-0.156
	Phase 2	0.927	311.656	2	49	.000	12011107.750	-62708.623	80.921
Lakshadweep	Phase 1	0.677	68.022	2	65	.000	-124.037	4.364	-0.016
	Phase 2	0.837	81.939	2	32	.000	-49073.519	677.552	-2.280
Puducherry	Phase 1	0.553	153.626	2	248	.000	-1320.115	33.095	-0.053
	Phase 2	0.996	7233.410	2	62	.000	750201.624	-4059.235	5.493
Tamil Nadu	Phase 1	0.745	382.982	2	262	.000	-24086.239	804.052	-2.379
	Phase 2	0.984	1852.479	2	60	.000	7462955.926	-39728.715	52.911
Telangana	Phase 1	0.747	378.885	2	256	.000	-6647.846	165.609	-0.124
	Phase 2	0.969	1010.937	2	65	.000	4814860.731	-25517.422	33.789



(a)



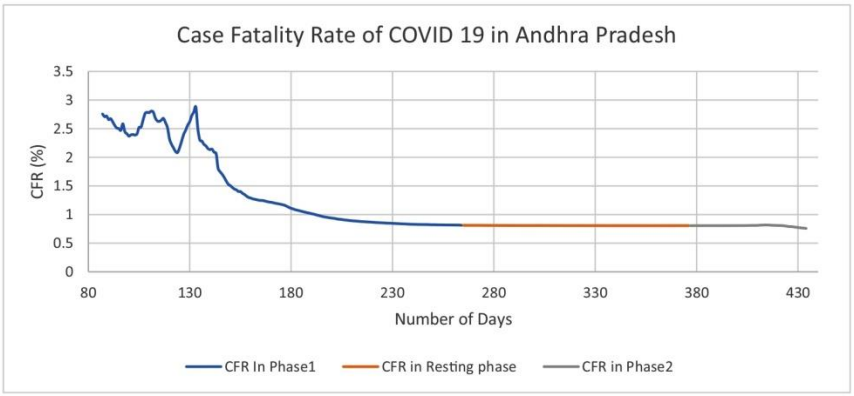
(b)



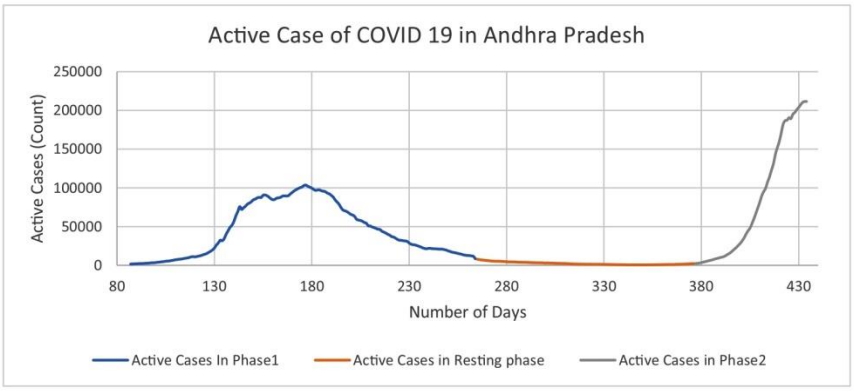
(c)

Figure 1. Andaman and Nicobar Islands

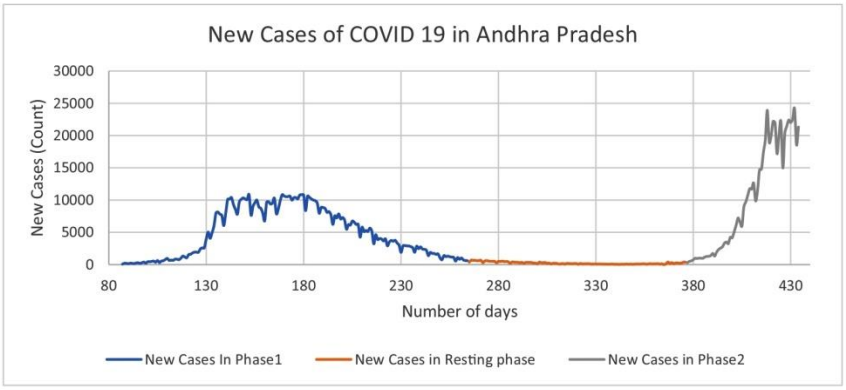
4



(a)



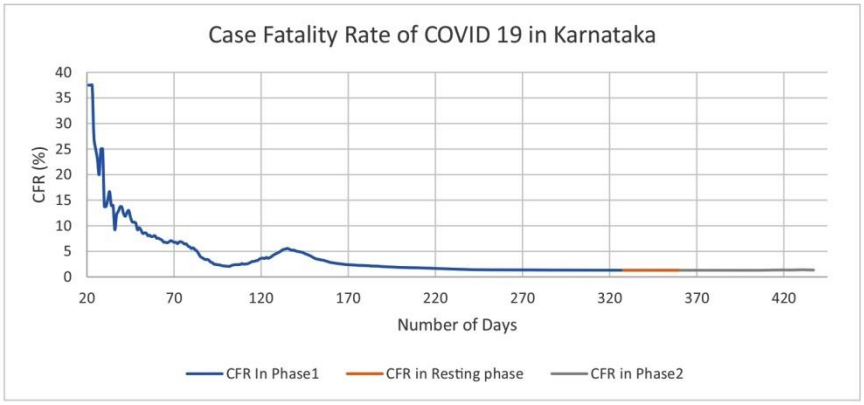
(b)



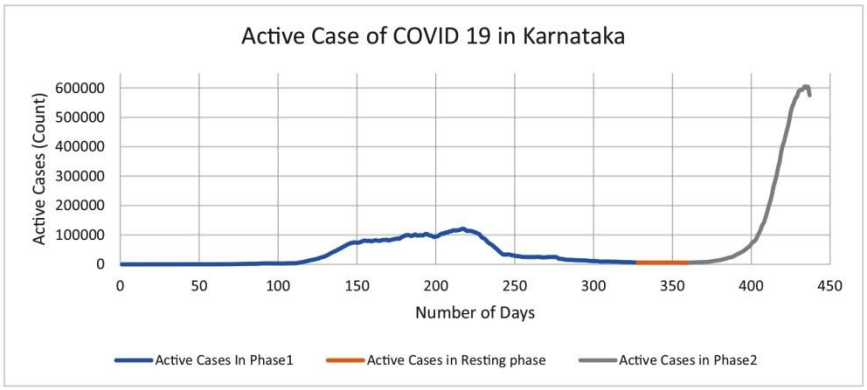
(c)

Figure 2. Andhra Pradesh

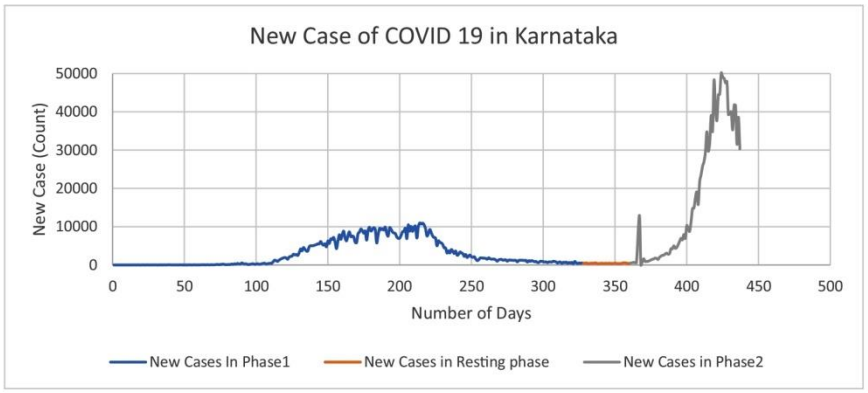
Similar to Karnataka (Figures 3b, 3c), Kerala (Figures 4b, 4c), Puducherry (Figures 6b, 6c), Tamil Nadu (Figures 7b, 7c), Telangana (Figures 8b, 8c) also had a 4.5 to 5 fold increase in the active and new cases when comparing phase one and two. The first state to report COVID cases in India was Kerala. However, the initial phase of Kerala had a long and stable initial quarter, and the increase has been triggered from the 200th day and continued towards 200 days. Whereas, Puducherry, Tamil Nadu, and Telangana initiated the trigger around the 100th day of the initial case reported. There were only 13 days in the resting phase in Kerala and 112, 63, and 97 respectively for Puducherry, Tamil Nadu, and Telangana. Thereafter the second phase had a very drastic and went up to 4.5 to 5 times for both Active Cases and New Cases.



(a)



(b)



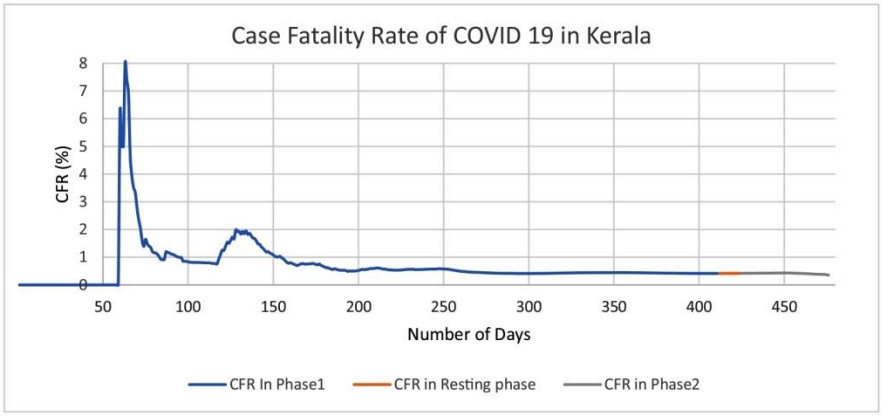
(c)

Figure 3. Karnataka

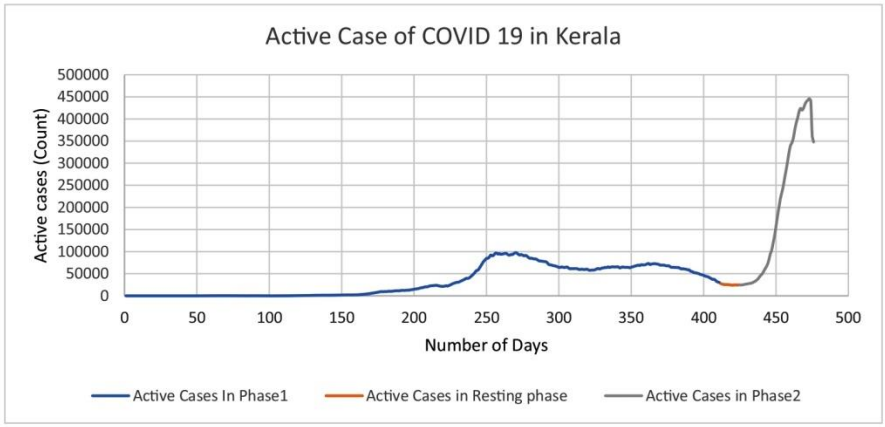
Discussion

COVID 19 is transmitted from the reservoir/source of infection to the susceptible host. The COVID 19 infection model is depicted in Figure 9. The reservoir, modalities of transmission, and susceptible host are the three main components of the transmission chain.[19]

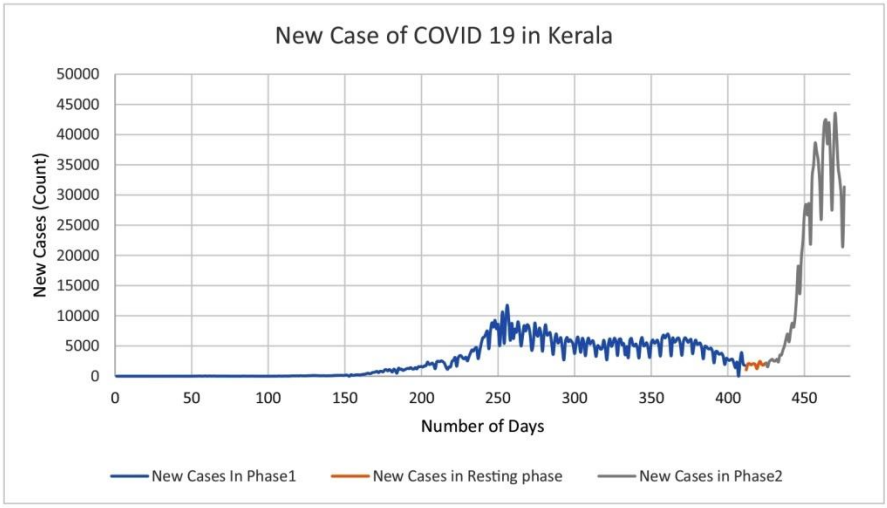
In phase one, the source or the reservoir was only the individuals who were infected and was travelling, whereas during the phase on, the transmission rate was very high, and the disease became community-based thereafter, the infection came to a resting phase. In the second phase, the source or the reservoir was almost many from the community itself therefore, a four to fivefold increase in the transmission within a very small



(a)

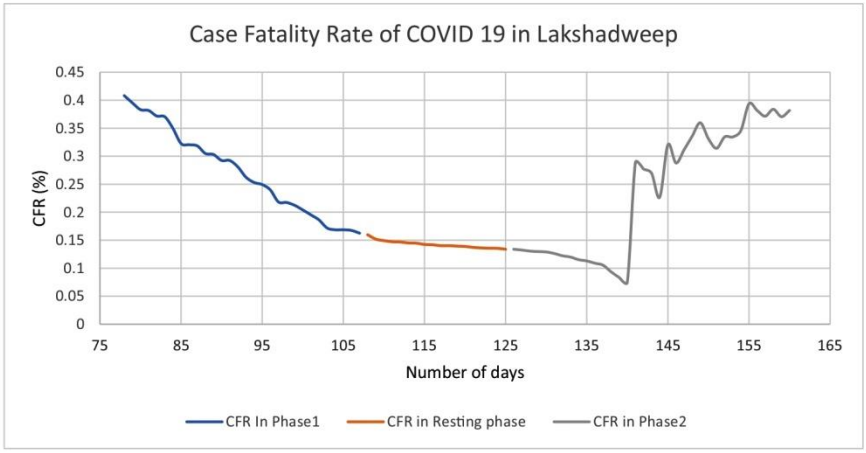


(b)

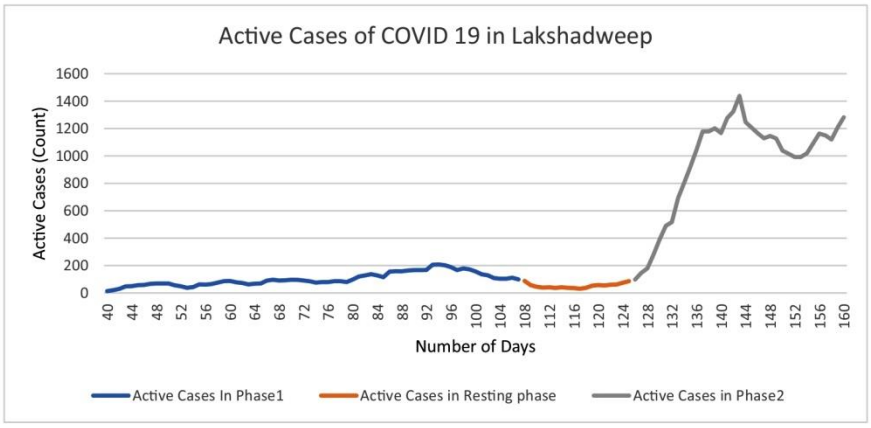


(c)

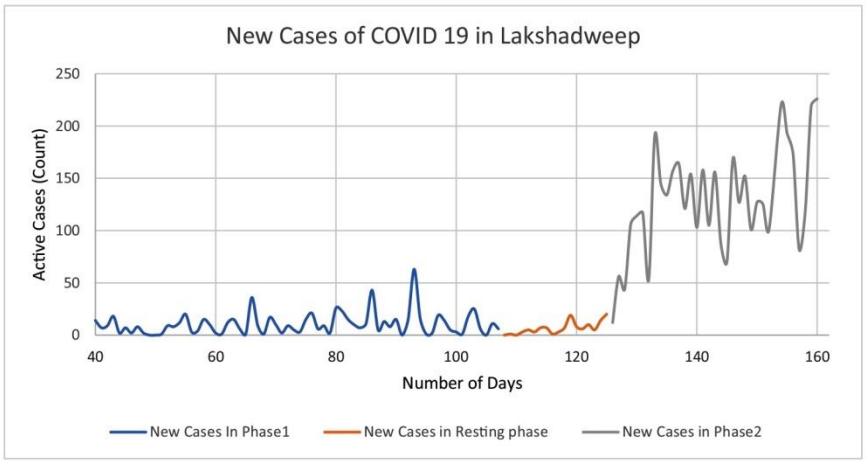
Figure 4. Kerala



(a)



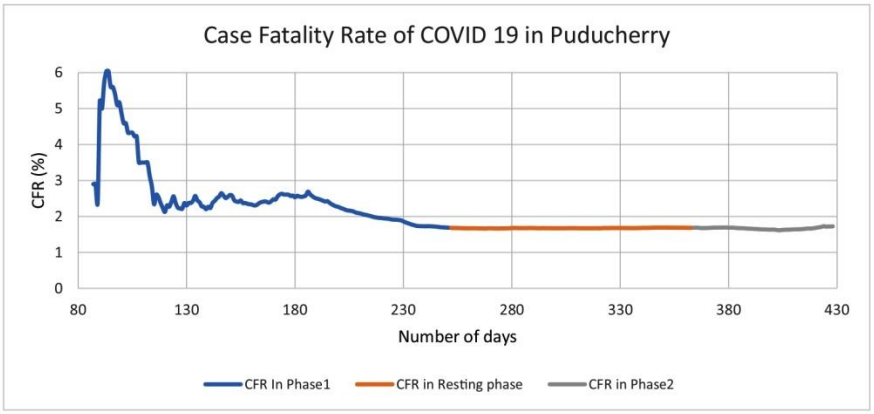
(b)



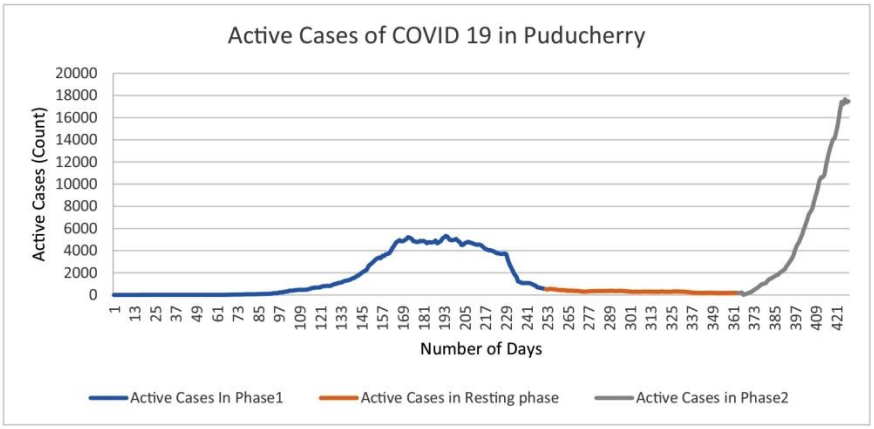
(c)

Figure 5. Lakshadweep

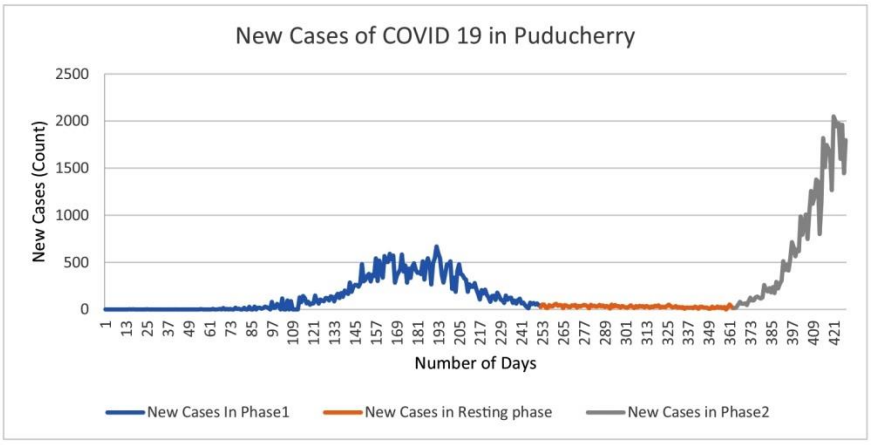
8



(a)



(b)

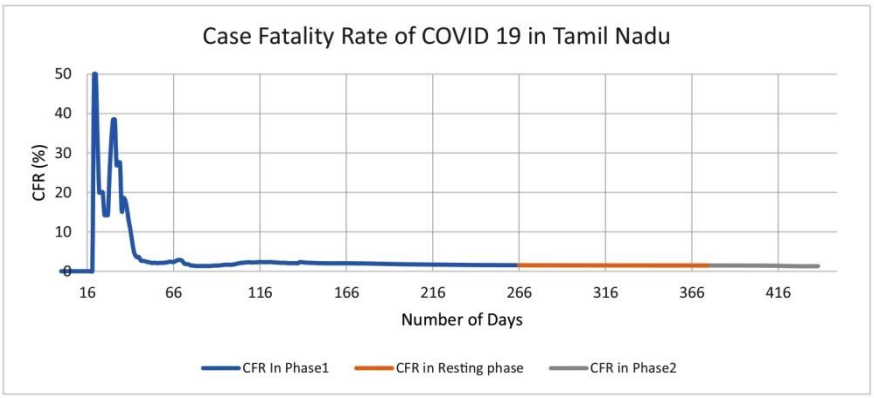


(c)

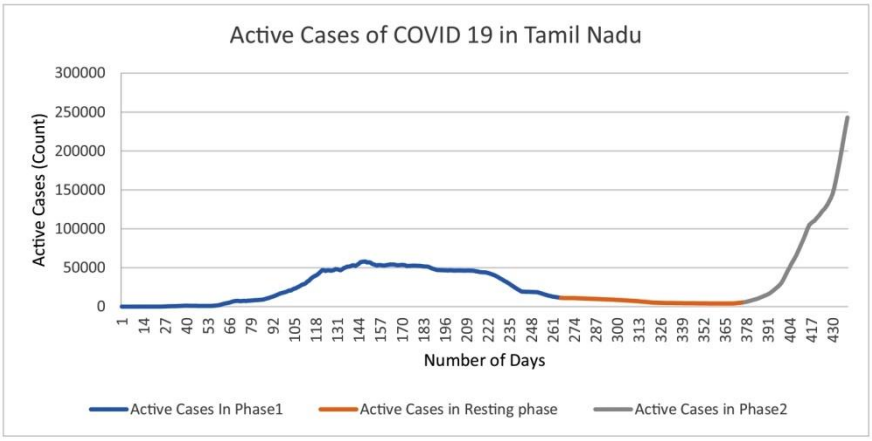
Figure 6. Puducherry

period of time. It was very clear that the second phase was very short but had very high intensity.

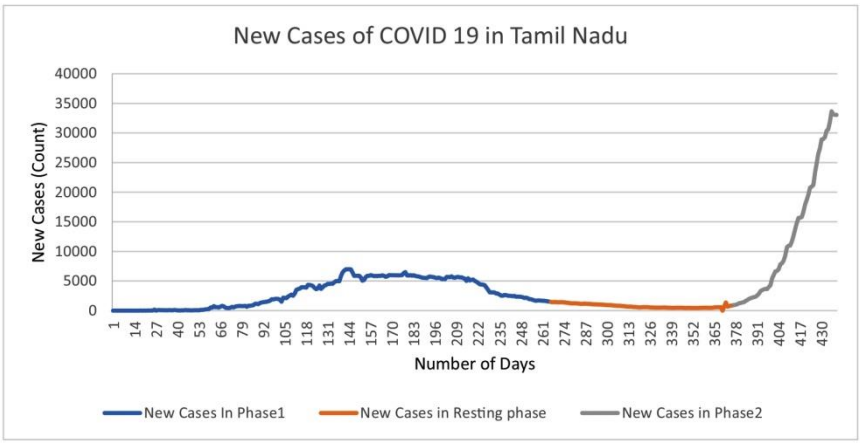
Kerala and Karnataka were having the highest days in phase one with 406 and 307 days, followed by Tamil Nadu and Telangana. Whereas the resting phase was more in Andra Pradesh, Tamil Nadu, and Telangana with 112, 111, and 97 days respectively, the state of Kerala and Karnataka was having the least with 13 and 33 days, respectively. When comparing the average CFR value in phase one was a



(a)



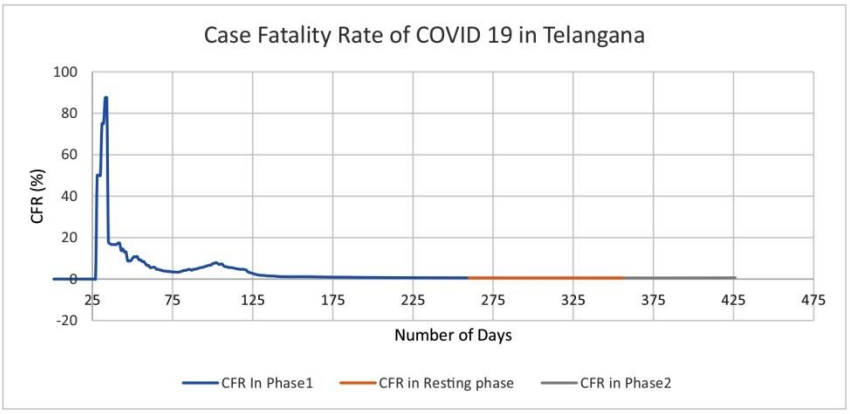
(b)



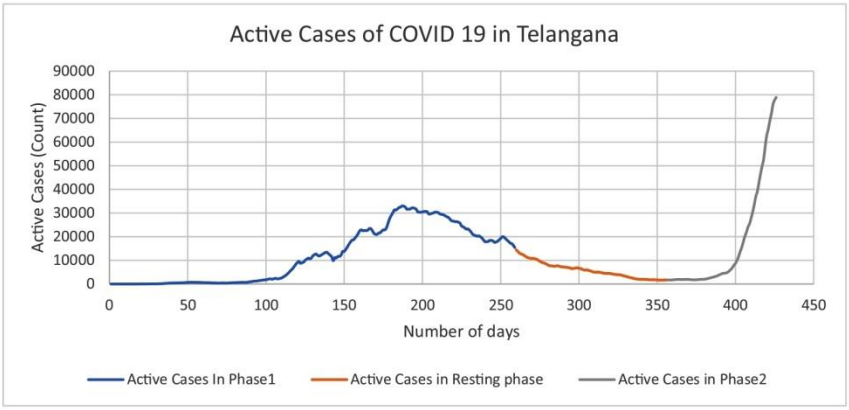
(c)

Figure 7. Tamil Nadu

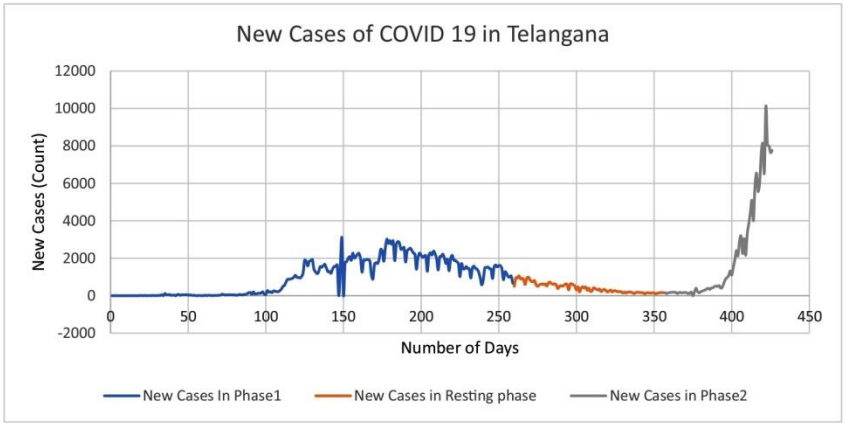
little higher as compared to the other phases in all of the south Indian states, Telangana Karnataka and Tamil Nadu was having higher CFR in phase one with 4.77, 4.22, and 3.71 respectively. At the same time, Lakshadweep and Kerala had the lowest with CFR of 0.27 and 0.71. When it comes to the resting phase the CFR stabilized in almost all the states and attained value between 0.2 and 2. Lakshadweep (0.143), Kerala (0.416), Telangana (0.553) and Andra Pradesh (0.803) had a very less CFR and Andaman and Nicobar Islands (1.237), Karnataka (1.306) and Tamil Nadu (1.490) had a comparatively high CFR. The trend was found in Phase Two also.



(a)



(b)



(c)

Figure 8. Telangana

When looking at the entire India, Rajesh and et.al. discovered that the CFR curve indicates a declining tendency, from 3.5 percent in mid-April 2020 to 1.2 percent in mid-April 2021.[20] Interestingly, while the virus appears to be more infectious in the second phase, the fall in the CFR curve suggests a silver lining in the form of a mutant that is less lethal. However, given the rapid increase in cases, it is projected that healthcare facilities will soon be fully throttled.

The paper also evaluated the optimal change points at which the COVID-19 cases turned their growth from linear to quadratic and then

Figure 9. Model of COVID 19 Infection

back to linear. Those periods initial to the optimal change was mentioned as the Phase one and the linear phase was known as Resting phase, and the period after the second cut off was identified as Phase two. The quadratic Regression Model was fitted and found to be the best model in each of the phases except the resting phase. The initial phase (phase one) was comparatively less fitted when compared to the later phase (phase two). The Quadratic regression model could also be used to predict the upcoming phases of the pandemic, which will have both the resting and non-resting periods. The Global maxima of the model were found to be around five times more in phase two when compared with phase one. The findings could be beneficial in influencing health policy decisions or government initiatives, especially developing initiatives aimed at reducing case fatality rates in their region over time, hence improving the overall health care facilities. These findings should, however, be seen to pave ways to more advanced mathematical and epidemiological models in research.

Conclusion

The first and second COVID-19 wave in India had different characteristics, the second which have been roughly said to be initialized on February 11, 2021 and was different in different states. The second phase had almost five time more when compared with the first and has been disastrous, with a number of cases exceeding 0.2 million per day on April 14, 2021. COVID-19 virus has over 7,684 varieties in India. As of May, 2021, the resurgent wave of COVID-19 in India has been spreading to rural population. The pandemic has brought attention to India's long-standing need for universal health coverage (UHC). Yet in the present moment, India confronts an urgent need to save lives and alleviate suffering.

The characteristics of the phases identified can be used for handling the upcoming waves. A one-size-fits-all approach is untenable since the numbers of COVID-19 cases and health services differ substantially from district to district in the upcoming waves. Working groups at the district level that have the autonomy to respond to quickly changing local situations must be empowered to receive resources and funds to coordinate efforts across the multifarious sectors of the health care system, i.e. from the front-line workers to tertiary care. Relevant technologies along with the statistical modelling could have a role in streamlining the management of resources and commodities, such as oxygen, ambulances, hospital beds, and funeral services.

Ethical Approval and Consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of supporting data

From
https://www.mohfw.gov.in/ through Kaggle.

Competing interests

The authors declare that they have no conflict of interest.

Funding

Not received any funding.

References

[1] Zhu H, Wei L, Niu P. The novel coronavirus outbreak in Wuhan. China glob health res policy. 2020;5:6. https://doi.org/10.1186/s41256-020-00135-6.
[2] Responding to COVID-19 - Learnings from Kerala. WHO. https://www.who.int/india/news/feature-stories/detail/responding-to-covid-19—learnings-from-kerala; 2020.
[3] Aiyar Yamini, Chandru Vijay, Chatterjee Mirai, Desai Sapna, Fernandez Armida, Gupta Atul, et al. K Sujatha Rao, Sharad Sharma, Devi Shetty, S V Subramanian, Leila E Caleb Varkey, Sandhya Venkateswaran, Vikram Patel, India's resurgence of COVID-19: urgent actions needed. The Lancet. 2021;397(10291):2232–4. ISSN 0140-6736. https://doi.org/10.1016/S0140-6736(21)01202-2.
[4] Kavitha C, Gowrisankar A, Banerjee S. The second and third waves in India: when will the pandemic be culminated? Eur Phys J Plus. 2021;136(5):596. https://doi.org/10.1140/epjp/s13360-021-01586-7. Epub 2021 May 28. PMID: 34094795; PMCID: PMC8163365.
[5] Sujita Kumar Kar, Ramdas Ransing S.M.Yasir Arafat, Vikas Menon, Second wave of COVID-19 pandemic in India: Barriers to effective governmental response, EClinicalMedicine, Volume 36, 2021, 100915, ISSN 2589-5370, doi: https://doi.org/10.1016/j.eclinm.2021.100915.
[6] Characterization of the Second Wave of COVID-19 in India Rajesh Ranjan, Aryan Sharma, Mahendra K. Verma medRxiv 2021.04.17.21255665; doi: https://doi.org/10.1101/2021.04.17.21255665
[7] Srivastava S, Banu S, Singh P, Sowpati DT, Mishra RK. SARS-CoV-2 genomics: An Indian perspective on sequencing viral variants. J Biosci. 2021;46(1):22. https://doi.org/10.1007/s12038-021-00145-7. PMID: 33737495; PMCID: PMC7895735.
[8] Kar SK, Ransing R, Arafat SMY, Menon V. Second wave of COVID-19 pandemic in India: Barriers to effective governmental response. EClinicalMedicine. 2021 Jun;36:100915. https://doi.org/10.1016/j.eclinm.2021.100915. Epub 2021 May 30. PMID: 34095794; PMCID: PMC8164526.
[9] Kumar SU, Kumar DT, Christopher BP, Doss CGP. The Rise and Impact of COVID-19 in India. Front Med (Lausanne). 2020 May;22(7):250. https://doi.org/10.3389/fmed.2020.00250. PMID: 32574338; PMCID: PMC7256162.
[10] Kuppalli Krutika, Gala Pooja, Cherabuddi Kartikeya. S P Kalantri, Manoj Mohanan, Bhramar Mukherjee, Lancelot Pinto, Manu Prakash, C S Pramesh, Sahaj Rathi, Nitika Pant Pai, Gavin Yamey, Madhukar Pai, India's COVID-19 crisis: a call for international action. The Lancet. 2021;397(10290):2132–5. ISSN 0140-6736. https://doi.org/10.1016/S0140-6736(21)01121-1.
[11] Noel George, Naresh K. Tyagi, Jang Bahadur Prasad, COVID-19 pandemic and its average recovery time in Indian states, Clinical Epidemiology and Global Health, Volume 11, 2021, 100740, ISSN 2213-3984, doi: https://doi.org/10.1016/j.cegh.2021.100740.
[12] Ghosh A, Roy S, Mondal H, Biswas S, Bose R. Mathematical modelling for decision making of lockdown during COVID-19 published online ahead of print, 2021 May 10. Applied Intelligence. 2021:1–17. https://doi.org/10.1007/s10489-021-02463-7.
[13] Pastor-Satorras, Romualdo, et al. "Epidemic Processes in Complex Networks." Reviews of Modern Physics, no. 3, American Physical Society (APS), Aug. 2015, Crossref, doi: https://doi.org/10.1103/revmodphys.87.925.
[14] Ananthakrishna, G. and Kumar, J., 2020. A reductive analysis of a compartmental model for COVID-19: data assimilation and forecasting for the United Kingdom. medRxiv preprint doi: https://doi.org/10.1101/2020.05.27.20114868
[15] Ansari A.A et al. 2021. Prevalence and cross states comparison of case fatality rate and recovery rate of COVID-19/SARS-COC-2 in India. 2021 Jan 10;1:475-480. doi: https://doi.org/10.4103/jfmpc.jfmpc_1088_20
[16] Kumar SU, Kumar DT, Christopher BP, Doss CGP. The Rise and Impact of COVID-19 in India. Front Med (Lausanne). 2020 May;22(7):250. https://doi.org/10.3389/fmed.2020.00250. PMID: 32574338; PMCID: PMC7256162.
[17] COVID-19 in India. (n.d.). Kaggle: Your Machine Learning and Data Science Community. https://www.kaggle.com/sudalairajkumar/covid19-in-india
[18] Gupta S, Kapoor V. Fundamentals of mathematical statistics. Sultan Chand & Sons. 2020. Twelth Edition, Page number 19.39.
[19] Park K. Park's textbook of preventive and social medicine; 2017.
[20] Ranjan R, Sharma A, Verma M. Characterization of the Second Wave of COVID-19 in India; 2021.

Impact of Lockdown on Air Quality in the Most Polluted Cities of

ijcm_980_22_R2_OA

Original Article

Impact of Lockdown on Air Quality in the Most Polluted Cities of India

Noel George, Jang Bahadur Prasad, Elizabeth Varghese[1], Richu Rajesh[2], Arvind Kumar[3]

Department of Epidemiology and Biostatistics, KLE University, Belgaum, Karnataka, [1]Cavinkare Private Limited, Chennai, Tamil Nadu, [2]Cochin University of Science and Technology, Kochi, Kerala, [3]School of Disaster Management, Tata Institute of Social Sciences, Mumbai, Maharashtra, India

Abstract

Background: COVID-19 has become a global pandemic, prompting lockdowns in practically every country. To prevent the spread of the disease, India has enforced a rigorous nationwide lockdown that commenced in March 2020. The lockdown imposed amid the pandemic ensured that most commercial activities and vehicle transportation ceased, resulting in a significant reduction in air pollution levels. **Methods:** The value of air pollutants PM10, PM2.5, NO_2, and SO_2 from January to May 2020 was obtained from the Indian Central Pollution Control Board. Before lockdown and during lockdown, relative fluctuations in ambient concentrations of four air contaminants were investigated. The Box–Jenkins approach was used to estimate future air pollution data points using time series data analysis. **Results:** The PM10 level reduced by 61%, 30%, 68%, 37%, and 43% in the selected cities, respectively. Comparison of other pollutant concentrations before and after the lockdown also found a reduction in ambient pollutant concentrations, resulting in improved air quality. Inference of predicted model values to observed values revealed a significant increase in the concentrations of all pollutants. The percentage increases in AQI_{mean} from predicted to observed values were 206% in Ghaziabad, 148% in Delhi, 59% in Hyderabad, and 160% in Cochin. **Conclusion:** The strict lockdown has resulted in a significant drop in air pollutant levels. Upgrading present technologies could help keep pollution to a minimum of 37% under control. The findings would prompt the government to consider how to strictly reduce vehicle and industrial pollution to improve air quality and maintain improved public health.

Keywords: Air pollutants, air quality indices, COVID-19, India, particulate matters

Introduction

Air pollution can be a complex mixture of gases, water vapor, particulate contaminants, and aerosols that have been expelled by human development and other natural/anthropogenic activities. Air pollution has been increasingly posing a threat to public health. It prompted the nation to take steps at various levels to come up with practical measures to minimize the damage. Science and technology plays an essential role in these efforts to reduce air pollution.[1] Different researchers, including government and non-governmental organizations, are trying to reduce the effects of air pollution by creating pollution-reducing devices and mechanisms.

Several countries around the world have been on lockdown as a result of a dramatic surge in the number of positive cases and deaths linked to COVID-19. As a result of the series of lockdowns, industrial activity, transportation by all modes, and virtually all other polluting activities have decreased considerably.[2] Almost every country experienced a significant reduction in automobile traffic and industrial activity, with the result being the cleanest and finest air quality in recent history. Lockdown events, however, factoring for meteorological changes lowered air pollutants, including weighted nitrogen dioxide and particulate matter levels, by nearly 60%, 31%, and 34%, respectively, with mixed impacts on ozone.[3] The decrease in NO_2 could be due to a decline in transportation sector emissions. The lockdown imposed in the aftermath of the COVID-19 pandemic ensured that most commercial activities ceased, resulting in a drastic reduction in air pollution levels.[4]

Access this article online

Quick Response Code:

Website: www.ijcm.org.in

DOI: 10.4103/ijcm.ijcm_980_22

Address for correspondence: Noel George, Department of Epidemiology and Biostatistics, KLE University, Belgaum - 590 010, Karnataka, India.
E-mail: noelgeorge2007@gmail.com

How to cite this article: George N, Varghese E, Rajesh R, Prasad JB, Kumar A. Impact of lockdown on air quality in the most polluted cities of India. Indian J Community Med 2023;XX:XX-XX.

Received: 12-12-22, **Revision:** ???,
Accepted: 28-10-23, **Published:** ***

Although major air pollutants such as PM10, PM2.5, carbon monoxide (CO), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), and ammonia (NH_3) showed significant reductions in concentrations during the lockdown period, and ozone (O_3) levels climbed in many places of the world.[5] The drop in PM2.5, CO, and NO_2 could also be due to reduced traffic and restricted industrial operations during the lockdown.[6] A study conducted in Delhi and Kolkata provides evidence of the reduction of PM10, PM2.5, NO_2, and CO concentrations, and the air quality in Delhi has improved by a maximum of 60%.[2,7] Therefore, it is proposed to model the optimal emission of pollutants by utilizing the pollutant emission data during the lockdown.

An air quality standard is a description of an enforceable level of air quality set by a regulatory authority. This paper focused on four parameters from the new Indian National Air Quality Standards (INAQS). According to the World Health Organization (WHO) database, India has 13 of the world's 20 most polluted cities[3] As per the World Air Quality Report, 2019, Ghaziabad and Delhi are the most polluted cities in India, whereas Kolkata and Hyderabad are among India's 10 most populated cities. This paper has considered Ghaziabad, Delhi, Kolkata, Hyderabad, and Cochin for the comparison of the impact of lockdown on air quality. Kerala, which is less polluted than the rest of the states, and the city of Cochin were chosen to learn more about the pollution trend in such areas.[8] The paper also tries to propose a control guideline for the emission of air pollutants in the upcoming period after lockdowns.

Materials and Methods

The secondary data were obtained from the Central Pollution Control Board through their website https://app.cpcbccr.com from January 2020 to May 2020. The 24-h average of each pollutant was used to compute the corresponding per-day levels. The data included the value of air pollutants PM_{10} (particulate matter less than 10 μm in diameter), $PM_{2.5}$ (particulate matter less than 2.5 μm in diameter), nitrogen dioxide (NO_2), and sulfur dioxide (SO_2). The Indian cities in the first 10 ranking of the World Air Quality Report, 2019, were taken into consideration except for Cochin. Cochin has been selected to know the trend in Kerala, which is lesser in pollution compared to others.

The monitoring of ambient air quality was carried out using different air quality indices (AQI). AQI can be defined as an overall value that converts weighted concentrations of individual air pollution-related parameters (particulate matter, SO_2, NO_2, etc.) into a single number or set of numbers.[2] This toxicity of the air was evaluated by classifying the AQIs based on a predefined criterion[9] [Table 1]. The table defines the guideline values for each pollutant.

Several types of AQIs with varying purposes and scopes have been developed in the recent past. Four different methods of estimating the AQI based on average pollutant concentration were used to compare the prevailing ambient air quality in the study regions. The study is performed in two compartments: before lockdown and during the lockdown. The time series method and its predicted value are then compared with the real-time values to ensure that the model is capable of predicting concentrations of different pollutants.

Four air quality approaches were illustrated in this study to estimate AQIs for two time periods for each city individually.

Method 1(AQI_{mean}): Ambient AQI was estimated by taking the arithmetic mean of the sum of the ratios of the four pollutants to their standard air quality values and was multiplied by 100. The air quality rating of each pollutant was estimated for AQI using the formula:

$$Q = \left(\frac{C}{Cs}\right) * 100 \quad (1)$$

where C is the observed value of the air quality pollutants and Cs is the CPCB standard for the given area (B, 2020a).

Method 2 (AQI_{gm})**:** Ambient AQI was estimated by taking the geometric mean of the AQIs provided by individual components (as in method 1) and was multiplied by 100. This measure is also compared with the quality scale provided by CPCB.

Method 3 (ORNAQI)**:** The Oak Ridge National Air Quality Index (ORNAQI), developed by the Oak Ridge National Laboratory (ORNL), USA, was estimated using the mathematical formula

$$AQI = [39.02 \sum \frac{Xi}{Xs}]^{0.967} \quad (2)$$

where X_i is the value of individual air quality parameters and X_s is the prescribed standard value for that parameters.[10] AQI measured by this method was then compared with relative ORAQI values.

Method 4 (AQI_{WeiAv})**:** This AQI was obtained by combining qualitative measures with the qualitative concept of the environment. The individual AQI is estimated as:

$$Q = \frac{W * C}{Cs} \quad (3)$$

where W is the weightage of the pollutant, C is the observed value of the pollutant, and Cs is the CPCB standard of pollutant

Table 1: Air quality ranges for AQIs

AQI (CPCB 2015)	ORNAQI	$AQI_{Wei\ Av}$
0-50	0≥ AQI ≤25	0.0≥ AQI ≤0.5
Good	Clean air	Acceptable
51-100	26≥ AQI ≤50	0.51≥ AQI ≤1.0
Satisfactory	Light air pollution	Unacceptable
101-200	51≥ AQI ≤75	1.01≥ AQI ≤2.0
Moderately polluted	Moderate	Alert
201-300	76≥ AQI ≤100	AQI ≥2.01
Poor	Heavy air pollution	Significantly harmful
301-400	AQI >101	
Very poor	Severe air pollution	
>401		
Severe		

for the given area. Here, all individuals are given equal weight (W = 1) and the total index is obtained as:

$$AQI = \sqrt{(\frac{1}{N})\Sigma_{i=1}^{N} Q_i^2} \quad (4)$$

where N is the number of air quality variables.[2]

A higher index value indicates higher levels of air pollution and, consequently, higher health risks. Forecasting of future data points of air pollutants 10 days posterior to the date was performed by the Box–Jenkins approach to time series Analysis.[11] The autoregressive integrated moving average model (ARIMA) was obtained by combining the autoregressive (AR) model and the moving average (MA) model.[10] The stationarity of data was evaluated using the Augmented Dicky Fuller test.[12] The time-series data after differencing is said to follow ARIMA (p, d, q), where parameters p is the lag order, d is the number of differencing and q is the order of moving average.[13]

By considering d = 1, we can obtain ARIMA (p, 1, q) with $\Delta_d y_t = W_t$ or $W_t = y_t - y_{t-1}$. Then,

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \ldots + \phi_p W_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \ldots - \theta_q \epsilon_{t-q}. \quad (5)$$

The data were analyzed using R version 4.0.0. A paired *t*-test was used to test the significant difference between two time periods (pre- and during lockdown) The Pearson correlation coefficient was used to identify the correlation between the individual air pollutants in the selected geographic areas. The level of $P > 0.05$ was considered significant.

Results

The CPCB data were used to calculate the average concentrations of ambient air pollutants PM10, PM2.5, SO_2, and NO_2, as well as the corresponding air pollution index. Some of the main AQIs were estimated for these periods before and during the lockdown. Table 2 shows the variations in air pollutant concentrations before and during lockdown periods. The reductions in the average concentrations of air pollutants in µgm/m^3 (PM_{10}, $PM_{2.5}$, SO_2, NO_2) were observed in Hyderabad (before: 99.33, 44.05, 27.92, and 9.52°, during: 62.04, 28.68, 16.92, and 8.41) and Cochin (before: 102.98, 16.51, 22.15, and 8.74°, during: 58.01, 9.82, 6.16 and 5.53) before and during lockdowns. Before the lockdown, pollutant concentrations were greater in Ghaziabad than in other cities, but during the lockdown, this highest position was switched to second, with Delhi being the city with the highest pollutant concentrations.

Variations in different AQI estimated for all five cities before and during the lockdown in India are shown in Table 3 and Figure 1.

In AQI_{mean}: Before the lockdown, the air in Ghaziabad, Delhi, Kolkata, Hyderabad, and Cochin was moderately polluted, satisfactory, satisfactory, and good, respectively, but at the time of the lockdown, there had been a noticeable change in Ghaziabad (satisfactory), Kolkata (good), and Hyderabad (good).

In AQI_{gm}: There was a drastic change in Ghaziabad, Delhi, and Kolkata from satisfactory to good air quality, whereas Hyderabad and Cochin already had good air quality with a reduction in the quantity of AQI values before and during the lockdown.

In ORNAQI: The quality of air in Ghaziabad has changed from severe to moderate air pollution, Delhi-severe to heavy air pollution, Kolkata-severe to light air pollution, Hyderabad and Cochin-moderate to light air pollution before and during the lockdown. But none of the cities fell under clean air range during the lockdown.

In AQI_{WeiAv}: The air quality in Ghaziabad and Delhi was at an alert level before the lockdown, which was changed to an unacceptable level during the lockdown. Similarly, Kolkata was at an alert level and has changed to an acceptable level during the lockdown. Hyderabad and Cochin were also changed from unacceptable to acceptable levels.

Table 4 shows the statistically significant difference in mean concentrations of ambient air pollutants in these cities over the comparison period. Except for the difference in

Table 2: Variation in the concentrations of pollutants in the selected cities of India

City	Time period	Average concentration of pollutants			
		PM_{10}	$PM_{2.5}$	NO_2	SO_2
Ghaziabad	Before lockdown	196.10	108.48	53.78	17.94
	During lockdown	101.82	41.26	20.49	17.54
Delhi	Before lockdown	178.80	107.70	41.45	15.90
	During lockdown	124.82	49.57	27.74	13.81
Kolkata	Before lockdown	181.201	96.17	36.50	9.07
	During lockdown	59.36	25.38	6.12	4.56
Hyderabad	Before lockdown	99.33	44.05	27.91	9.52
	During lockdown	62.04	28.68	16.92	8.41
Cochin	Before lockdown	102.98	16.51	22.15	8.74
	During lockdown	58.01	9.82	6.16	5.53

Table 3: AQI values of the selected cities of India

City	Time Period	AQI			
		AQI_{mean}	AQI_{gm}	ORNAQI	AQI_{WeiAv}
Ghaziabad	Before lockdown	115.43	82.22	174.45	1.39
	During lockdown	52.94	43.29	82.07	0.63
Delhi	Before lockdown	99.08	56.72	130.52	1.27
	During lockdown	60.30	40.23	80.85	0.76
Kolkata	Before lockdown	177.73	60.61	130.88	1.23
	During lockdown	45.39	17.39	39.49	0.37
Hyderabad	Before lockdown	54.88	40.55	73.87	0.64
	During lockdown	35.40	27.72	48.38	0.41
Cochin	Before lockdown	42.28	26.91	57.29	0.57
	During lockdown	22.25	14.64	30.88	0.30

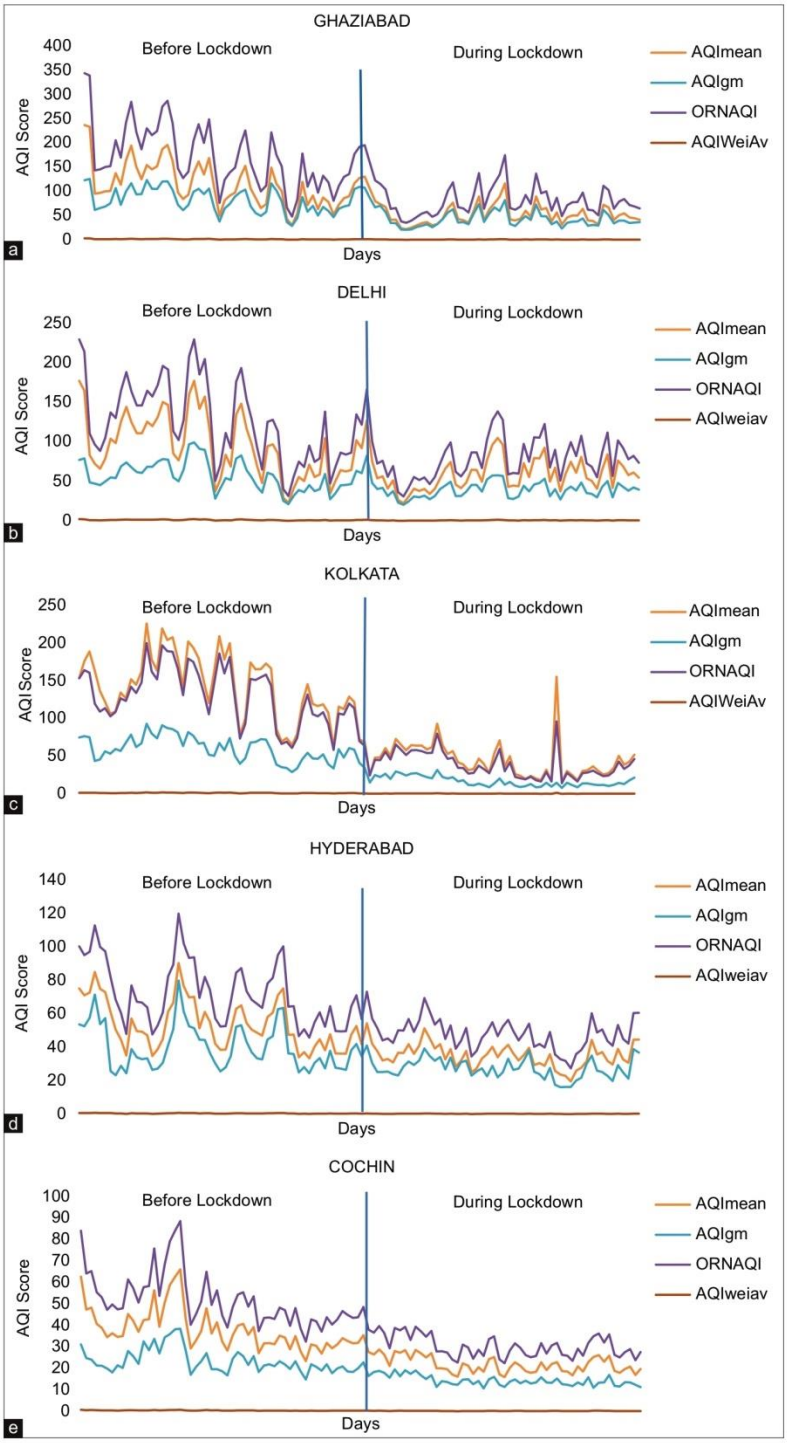


Figure 1: Variation in air quality before and during the lockdown in various Indian cities

SO_2 concentration at Ghaziabad and Delhi, the difference in the mean values of all parameters between the two time periods (before and after lockdown) was statistically significant ($P > 0.05$). The Pearson coefficient of correlation for cities during study periods is shown in Table 5. This shows that before the lockdown, cities had a positive correlation between the parameters PM_{10}, $PM_{2.5}$, SO_2, and NO_2, except for Cochin. During the pre-lockdown period, Ghaziabad, Delhi, and Kolkata showed a higher correlation between PM_{10} and $PM_{2.5}$. For Ghaziabad and Delhi, there was the least correlation between SO_2 and PM_{10} during both periods. However, in Delhi, during the lockdown period, the higher correlation remained the same, but the lowest correlation was between the pollutants NO_2 and PM_{10}. It should be noted that the deviation in the SO_2 level of Ghaziabad was not significant. Although in Hyderabad, there was a higher correlation between PM_{10} and SO_2 and a lower correlation between NO_2 and $PM_{2.5}$ before the lockdown and during the lockdown, there existed a higher correlation between NO_2 and SO_2 and a lower correlation between PM_{10} and NO_2. In the case of Cochin, the correlation between the pollutants was very weak, and the pollutants NO_2 and PM_{10} showed a negative correlation in the pre-lockdown period. During the lockdown, a negative correlation existed between SO_2 and PM_{10}. The results from the correlation coefficient stated that the strength of the correlation between the pollutants was reduced during the lockdown period.

It is shown that average PM_{10} levels are reduced by 61%, 30%, 68%, 37%, and 43%, respectively, in Ghaziabad, Delhi, Kolkata, Hyderabad, and Cochin. Similarly, in the case of $PM_{2.5}$, the average concentrations are reduced at a rate of 61%, 53%, 73%, 34%, and 40%, respectively. At Ghaziabad, Delhi, Kolkata, Hyderabad, and Cochin, NO_2 is reduced at 61%, 33%, 83%, 39%, and 72% respectively. The decrease in the concentrations of SO_2 is very low when compared to other

Table 4: Results of paired *t*-test for the mean difference in pollutants of the selected cities of India

City	Pollutants											
	PM_{10}			$PM_{2.5}$			NO_2			SO_2		
	Mean difference	95% CI	*P*	Mean difference	95% CI	*P*	Mean difference	95% CI	*P*	Mean difference	95% CI	*P*
Ghaziabad	94.28	69.96, 118.62	<0.01	67.55	50.04, 82.22	<0.01	33.29	27.10, 37.29	<0.01	00.36	-3.91, 2.78	0.73
Delhi	48.99	24.33, 73.65	<0.01	51.06	36.77, 65.35	<0.01	15.89	12.30, 19.47	<0.01	03.64	-2.88, 0.16	0.08
Kolkata	121.65	104.17, 139.12	<0.01	70.79	61.18, 180.40	<0.01	30.38	28.17, 35.68	<0.01	4.50	3.33, 5.67	<0.01
Hyderabad	37.50	29.734, 45.27	<0.01	13.14	9.11, 17.18	<0.01	14.56	8.41, 13.12	<0.01	01.35	0.88, 3.59	0.02
Cochin	38.29	32.38, 44.21	<0.01	06.76	2.17, 11.35	0.004	10.98	7.76, 14.20	<0.01	0.64	0.30, 0.98	<0.01

Table 5: Correlation between the pollutants PM_{10}, $PM_{2.5}$, SO_2, and NO_2 in the selected cities of India

City	Pollutant	Correlation between individual pollutants							
		Before lockdown				After lockdown			
		PM_{10}	$PM_{2.5}$	NO_2	SO_2	PM_{10}	$PM_{2.5}$	NO_2	SO_2
Ghaziabad	PM_{10}	1.000	0.957	0.725	0.153	1.000	0.895	0.709	0.418
	$PM_{2.5}$	0.957	1.000	0.669	0.618	0.895	1.000	0.720	0.532
	NO_2	0.725	0.669	1.000	0.288	0.709	0.720	1.000	0.390
	SO_2	0.153	0.618	0.288	1.000	0.418	0.532	0.390	1.000
New Delhi	PM_{10}	1.000	0.941	0.614	0.382	1.000	0.836	0.304	0.793
	$PM_{2.5}$	0.941	1.000	0.461	0.215	0.836	1.000	0.536	0.710
	NO_2	0.614	0.461	1.000	0.487	0.304	0.536	1.000	0.319
	SO_2	0.382	0.215	0.487	1.000	0.793	0.710	0.319	1.000
Kolkata	PM_{10}	1.000	0.747	0.558	0.089	1.000	0.375	0.236	0.230
	$PM_{2.5}$	0.747	1.000	0.604	0.465	0.375	1.000	0.524	0.559
	NO_2	0.558	0.604	1.000	0.100	0.236	0.524	1.000	0.695
	SO_2	0.089	0.129	0.100	1.000	0.230	0.559	0.695	1.000
Hyderabad	PM_{10}	1.000	0.465	0.647	0.785	1.000	0.349	0.190	0.608
	$PM_{2.5}$	0.615	1.000	0.494	0.652	0.349	1.000	0.293	0.269
	NO_2	0.647	0.494	1.000	0.777	0.190	0.293	1.000	0.723
	SO_2	0.785	0.652	0.777	1.000	0.608	0.269	0.723	1.000
Cochin	PM_{10}	1.000	0.007	-0.118	0.006	1.000	0.244	0.255	-0.123
	$PM_{2.5}$	0.007	1.000	0.160	0.338	0.244	1.000	0.796	0.175
	NO_2	-0.118	0.160	1.000	0.155	0.255	0.796	1.000	0.165
	SO_2	0.006	0.338	0.155	1.000	-0.123	0.175	0.165	1.000

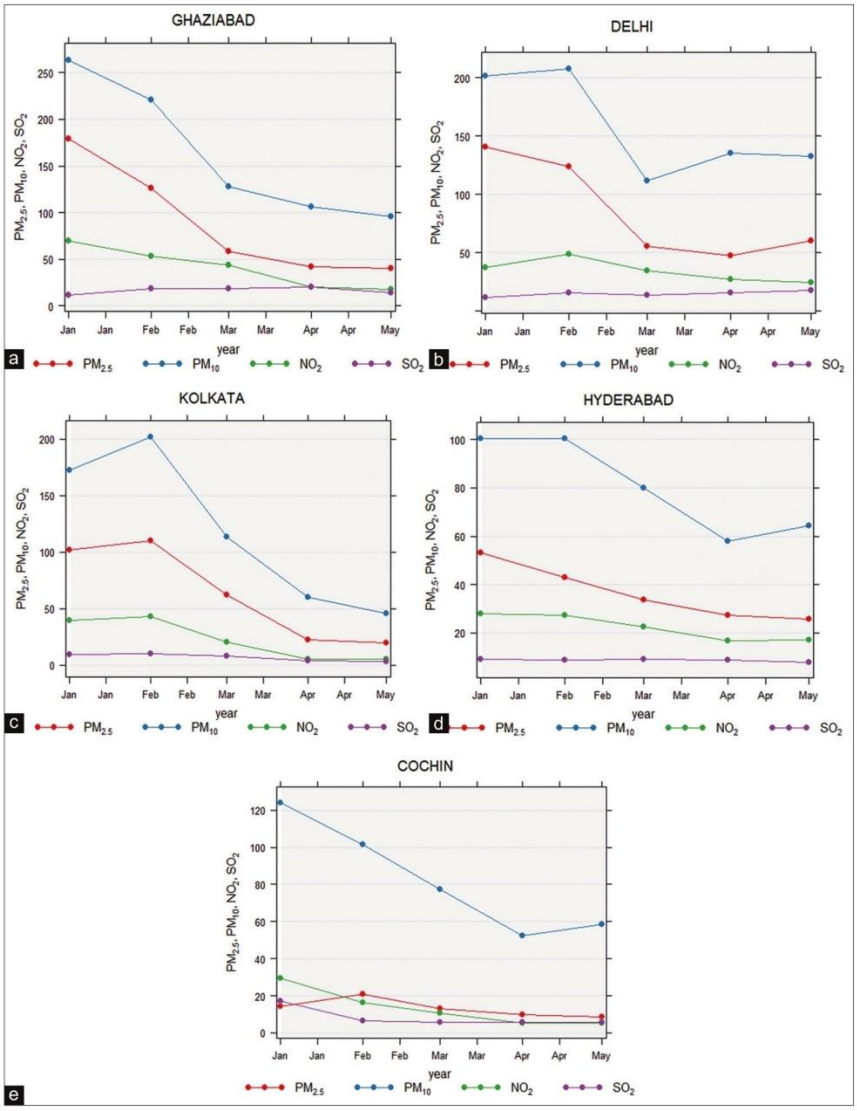


Figure 2: Trend in the variations in pollutant concentrations before and during the lockdown in various Indian cities

pollutants. SO_2 undergoes a decline of 2% at Ghaziabad, 13% at Delhi, 49% at Kolkata, 11% at Hyderabad, and 36% at Cochin.

The ARIMA model for all orders of p, d, and q was built, and chosen the best model, which has the least AIC value. Table 6 shows the selected ARIMA models with predicted values of the air pollutants and AQI_{mean}, for the 10th day after 12th May 2020 (May 22, 2020). Inference of predicted model values to observed values revealed a significant increase in the concentrations of all pollutants (excluding PM_{10} in Kolkata). Based on the predicted AQI_{mean}, the air quality in Ghaziabad and Delhi on May 22, 2020, was satisfactory, whereas Kolkata, Hyderabad, and Cochin were in the good air quality range. Observed values inferred that Ghaziabad and Delhi were in a moderately polluted range, Hyderabad was satisfactory, whereas Kolkata and Cochin remained good. The percentage increases in the AQI_{mean} from predicted to observed value were 206% in Ghaziabad, 148% in Delhi, 59% in Hyderabad, and 160% in Cochin, whereas in Kolkata AQI_{mean} predicted and observed values were 28.

Discussion

The current study shows that the COVID-19 pandemic-related lockdown has a direct impact on improving ambient air quality in Indian cities, as indicated by significant reductions in most air pollutants, particularly PM_{10}, $PM_{2.5}$, SO_2, and NO_2. These reduced levels of AQI are within the allowed limits set by the Indian National Pollution Control Agency, namely the Central Pollution Control Board (CPCB) regulations [Table 1].

Table 6: Selected ARIMA model, forecasted values and % change of pollutants (PM_{10}, $PM_{2.5}$, SO_2, and NO_2) and AQI_{mean} with observed value

City	PM_{10}			$PM_{2.5}$		
	Selected ARIMA	Predicted Value	Observed (% change)	Selected ARIMA	Predicted Value	Observed (% change)
Ghaziabad	(3,1,1)	95	201 (112)	(2,1,1)	40	194 (385)
Delhi	(2,1,1)	124	158 (27)	(2,1,2)	55	222 (304)
Kolkata	(2,1,2)	55	42 (-24)	(0,1,3)	27	32 (19)
Hyderabad	(3,1,3)	81	120 (48)	(3,1,3)	28	53 (89)
Cochin	(2,1,3)	10	22 (120)	(1,1,2)	10	20 (100)

City	NO_2			SO_2			AQI_{mean}	
	Selected ARIMA	Predicted Value	Observed (% change)	Selected ARIMA	Predicted Value	Observed (% change)	Predicted Value	Observed (% change)
Ghaziabad	(0,1,3)	19	85 (347)	(3,1,3)	14	31 (121)	51	156 (206)
Delhi	(2,1,1)	23	57 (148)	(3,1,1)	18	61 (239)	62	154 (148)
Kolkata	(3,1,2)	6	11 (83)	(1,1,2)	3	4 (33)	28	28 (0)
Hyderabad	(3,1,3)	23	38 (65)	(1,1,1)	9	16 (78)	44	70 (59)
Cochin	(0,1,3)	5	26 (420)	(2,1,3)	4	15 (275)	10	26 (160)

Before the lockdown began, Ghaziabad and Kolkata were the most polluted cities on the list. However, when the lockdown began, both cities experienced sudden changes in air quality compared to other cities. Among five of the selected cities, Cochin remained the least polluted in both periods. With the implementation of the lockdown, air pollution levels in India's most polluted cities met the CPCB ambient air quality criteria (Organization., 2016). The reasons for this reduction were the strict implementation of laws and regulations such as prohibiting all outdoor activities, closing transportation sectors, closing industries, markets, workplaces, offices, schools, colleges, and any other institutional areas framed to combat the disease outbreak in hotspot regions.[14]

To combat the spread of the coronavirus pandemic, a 21-day lockdown was imposed across the country from March 25, 2020. Except for critical services, all industries have been shut down and the government has implemented very strict restrictions, which were reflected in the concentration decline of air pollutants. With very few relaxations, the Government of India and the Ministry of Home Affairs (MHA) extended the lockdown duration to 2 weeks beyond May 4 on May 1.[15] The current study considered, predicting concentrations of different pollutants and AQI_{mean}, indicating the upcoming changes in the air quality as a result of lockdown measures. The concentration of pollutants and AQI during the pre-lockdown period were found to be at a very high level, whereas the concentration drastically changed to an almost acceptable level during the strict lockdown. The pre-lockdown period, which lasted until March 24, 2020, was the first. Following the first lockdown, three additional lockdown stages followed in quick succession (LD 2: April 15–May 3, 2020, LD 3: May 4–May 17, 2020, LD 4: May 18–May 31, 2020).[16] The National Disaster Management Authority announced on May 17 that the lockdown would be prolonged for another 2 weeks, with other additional relaxations in transportation and opening of shops. If a strict lockdown scenario existed until May 22, 2020, the air quality would be similar to the study's predicted values, but the additional relaxations, which were given two times after May 11, 2020, resulted in an increase in the observed concentrations and AQI_{mean} (except Kolkata) of pollutants compared to the predicted values on May 22, 2020. Due to the sharp increase in COVID-19 instances, the West Bengal government enforced near-complete lockdown regulations throughout the state, allowing only emergency movement of persons and vehicles. As a result of the decrease in contaminants, the AQI_{mean} predicted and observed were the same. Thus, the COVID-19 pandemic lockdown gave awareness to all governments throughout the world about the importance of restoring environmental quality and natural ecosystem stability.[17]

Predicting air quality is a useful strategy for air pollution management that the local government and municipality might employ in the future. It is beneficial to act quickly before the situation worsens in the long run. Thus, the study clearly shows that improved model performance is critical for accurate forecasting of air quality.

Conclusion

Air quality was significantly improved when the lockdown was imposed in Ghaziabad, Delhi, Kolkata, Hyderabad, and Cochin. A markable reduction was observed in the ambient concentration of PM_{10}, $PM_{2.5}$, SO_2, and NO_2 pollutants during the lockdown days of 2020. These current environmental gains are quite short, and this short-term improvement in air quality during lockdown can provide a promising signal to the governments and policymakers to improve the quality of air through a planned strict restriction on pollution sources. For a period of time, the government can consider a lockdown at the hotspot pollution areas to manage the level of pollution with

low economic loss. A further increase in the concentration of air pollutants can be handled if proper measures are taken at the right time.

Acknowledgments

The authors acknowledge the support provided by the staff and students of St. Thomas College, Palai, for their expertise and assistance throughout our study.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Manisalidis I, Stavropoulou E, Stavropoulos A, Bezirtzoglou E. Environmental and health impacts of air pollution: A review. Front Public Health 2020;8:14.
2. Wang Q, Su M. A preliminary assessment of the impact of COVID-19 on environment – A case study of China. Sci Total Environ 2020;728:138915.
3. World Health Organization. Ambient air pollution: A global assessment of exposure and burden of disease 2016. Available from: https://apps.who.int/iris/handle/10665/250141.
4. Significant improvement in air quality across India due to coronavirus lockdown: CPCB- The New Indian Express. (n.d.). Available from: https://www.newindianexpress.com/nation/2020/apr/02/significant-improvement-in-air-quality-across-india-due-to-coronavirus-lockdown-cpcb-2124967.html. [Retrieved 2021 Aug 15].
5. Sekar A, Jasna RS, Binoy BV, Mohan P, Varghese GK. Air quality change due to COVID-19 lockdown in India and its perception by public, 11 September 2020, PREPRINT (Version 1) available from: Research Square. doi: 10.21203/rs. 3.rs-74610/v1.
6. Misra P, Takigawa M, Khatri P, Dhaka SK, Dimri AP, Yamaji K, *et al*. Nitrogen oxides concentration and emission change detection during COVID-19 restrictions in North India. Sci Rep 2021;11:9800.
7. Bhat SA, Bashir O, Bilal M, Ishaq A, Din Dar MU, Kumar R, *et al*. Impact of COVID-related lockdowns on environmental and climate change scenarios. Environ Res 2021;195:110839.
8. World's Most Polluted Cities in 2020-PM2.5 Ranking | AirVisual. (n.d.). Retrieved August 15, 2021. Available from: https://www.iqair.com/world-most-polluted-cities.
9. Ravikumar P, Prakash KL, Somashekar RK. Air quality indices to understand the ambient air quality in vicinity of dam sites of different irrigation projects in Karnataka state, India. IJSN 2014;5:531-41.
10. Arltová M, Fedorova D. Selection of unit root test on the basis of length of the time series and value of AR (1) parameter. STATISTIKA 2016;3:47-96.
11. Lee MH, Rahman NHA, Suhartono, Latif MT, Nor ME, Kamisan NAB. Seasonal ARIMA for forecasting air pollution index: A case study. Am J Appl Scie 2012;9:570-8.
12. Zambrano-Monserrate MA, Ruano MA, Sanchez-Alcalde L. Indirect effects of COVID-19 on the environment. Sci Total Environ 2020;728:138813.
13. Montgomery DC, Jennings CL, Kulahci M. Introduction to Time Series Analysis and Forecasting. 2nd ed. Hoboken, New Jersey: John Wiley and Sons; 2005. p. 671.
14. Garg A, Kumar A, Gupta NC. Impact of lockdown on ambient air quality in COVID-19 affected hotspot cities of India: Need to readdress air pollution mitigation policies. Environ Claims J 2020;33:65-76.
15. Full list of Red, Yellow, Green Zone districts for Lockdown 3.0-India News. (n.d.). Retrieved September 9, 2021. Available from: https://www.indiatoday.in/india/story/red-orange-green-zones-full-current-update-list-districts-states-india-coronavirus-1673358-2020-05-01.
16. Soni P. Effects of COVID-19 lockdown phases in India: An atmospheric perspective. Environ Dev Sustain 2021;23:12044-55.
17. Coronavirus India lockdown Day 49 updates | May 12, 2020-The Hindu. (n.d.). Retrieved September 9, 2021. Available from: https://www.thehindu.com/news/national/india-coronavirus-lockdown-may-12-2020-live-updates/article31563054.ece.

ANNEXURE 2A- ORAL PRESENTATION CERTIFICATE

MLC FOUNDATION
'SHYAM' INSTITUTE

This is to certify that

Noel George

participated in

XXI Bhopal Seminar 2023

CONTEMPORARY ISSUES IN DEMOGRAPHY AND DEVELOPMENT IN INDIA

and presented the paper

COVID-19 pandemic and its average recovery time in Indian states

Aalok Ranjan
24 February 2023

www.mlcfoundation.org.in
www.shyaminstitute.in

XXVII IIPS NATIONAL SEMINAR 2023

CERTIFICATE

75 Years of India's Demographic Change: Processes and Consequences

This is to certify that ***Prof./Dr./Mr./Ms. Noel George*** has presented a paper/poster titled ***COVID-19 pandemic and its average recovery time in Indian states*** in the **XXVII IIPS National Seminar 2023** on **"75 years of India's Demographic Change: Processes and Consequences"** jointly organised by International Institute for Population Sciences (IIPS), Mumbai and Institute for Social and Economic Change (ISEC), Bengaluru conducted during February 23-25, 2023, at ISEC, Bengaluru.

D Rajasekhar
Director and HAG Professor
ISEC, Bengaluru

K S James
Director and Senior Professor
IIPS, Mumbai

The Department of Statistics
Pachhunga University College Campus
Mizoram University (A Central University), Aizawl

Awards this Certificate to

Noel George, KLE University

for presenting a paper entitled "**Statistical Model for COVID-19 in Different Waves of South Indian States**" in the **International Conference on Applications of Statistics**, organized by the Department of Statistics, Pachhunga University College (PUC) campus, Mizoram University (MZU), during 2-3 February, 2023.

Prof. H. Lalthanzara *Principal, PUC campus, MZU*	**Dr. Lalpawimawha** *Assistant Professor, Head, Dept. of Statistics, PUC campus, MZU*	**Dr. Samba Siva Rao Pasupuleti** *Assistant Professor, Dept. of Statistics, PUC campus, MZU (Organizing Secretary)*	**Dr. Mukesh Ranjan** *Assistant Professor, Dept. of Statistics, PUC campus, MZU (Organizing Secretary)*

ANNEXURE 2B- WHO CERTIFICATES

World Health Organization

HEALTH EMERGENCIES programme

Confirmation of Participation

Learning how to use the COVID-19 Vaccine Introduction and deployment Costing (CVIC) tool

noel george

November 29, 2023

World Health Organization

HEALTH EMERGENCIES programme

Record of Achievement

COVID-19: Operational Planning Guidelines and COVID-19 Partners Platform to support country preparedness and response

noel george

April 23, 2020

Verify online: https://openwho.org/verify/xilov-cozed-mipul-sutaz-hovyl

World Health Organization

HEALTH EMERGENCIES programme

Record of Achievement

Prevention, identification and management of infections in health workers in the context of COVID-19

noel george

April 12, 2022

Verify online: https://openwho.org/verify/xogod-hucag-sacin-sasik-hefym

World Health Organization

HEALTH EMERGENCIES programme

Certificate of Achievement

This is to certify that

noel george

successfully completed and received a passing grade in

Transmission-based Precautions

Verified certificate
issued on:
April 12, 2022

Verify online: https://openwho.org/verify/xocol-fumyp-bupis-celyc-vokih

a course offered on OpenWHO.org,
an online learning initiative of WHO Health Emergencies Programme.

World Health Organization

HEALTH EMERGENCIES programme

Record of Achievement

Guidance on mask use in the context of COVID-19

noel george

April 13, 2022

Verify online: https://openwho.org/verify/xoved-fasoh-todes-tadyz-lamuv

World Health Organization

HEALTH EMERGENCIES programme

Record of Achievement

COVID-19 and work:
Staying healthy and safe at work during the COVID-19 pandemic

noel george

April 14, 2022

Verify online: https://openwho.org/verify/xuteb-vovir-tamyt-capan-kogek

World Health Organization

HEALTH EMERGENCIES programme

Certificate of Achievement

This is to certify that

noel george

successfully completed and received a passing grade in

Leadership and programme management in Infection Prevention and Control (IPC)

Verified certificate
issued on:
April 15, 2022

Verify online: https://openwho.org/verify/xuvic-pylyg-zobys-vecod-pozyl

a course offered on OpenWHO.org,
an online learning initiative of WHO Health Emergencies Programme.

World Health Organization

HEALTH EMERGENCIES programme

Certificate of Achievement

This is to certify that

noel george

successfully completed and received a passing grade in

Clinical management of patients with COVID-19: Rehabilitation of patients with COVID-19

Verified certificate
issued on:
April 16, 2022

Verify online: https://openwho.org/verify/xerab-dyryp-kavic-mydat-titin

Certificate validity
1 year from date of issuance.

a course offered on OpenWHO.org,
an online learning initiative of WHO Health Emergencies Programme.

World Health Organization

HEALTH EMERGENCIES programme

Record of Achievement

The Public Health Emergency Operations Centre (PHEOC)

noel george

April 16, 2022

CPD CERTIFIED
The CPD Certification Service

Verify online: https://openwho.org/verify/xeded-rytis-dehed-vyrot-kazeh

World Health Organization

HEALTH EMERGENCIES programme

Record of Achievement

SARS-CoV-2 antigen rapid diagnostic testing

noel george

April 17, 2022

Verify online: https://openwho.org/verify/xilig-lycub-vofic-pelal-gokez

World Health Organization

HEALTH EMERGENCIES programme

Record of Achievement

Learning package for Rapid Response Teams in the context of COVID-19 in India

noel george

April 17, 2022

Verify online: https://openwho.org/verify/xufeg-pidep-mosem-dulap-pekur

World Health Organization

HEALTH EMERGENCIES programme

Record of Achievement

Clinical management of patients with COVID-19 Investigations and care for mild, moderate and severe disease

noel george

April 18, 2022

Verify online: https://openwho.org/verify/xuhez-ruzyb-hotod-gerul-kupah

World Health Organization

Vaccine Safety Basics

Record of Achievement

Vaccine Safety Basics

noel george

April 18, 2022

Verify online: https://openwho.org/verify/xovoc-gumok-kivuf-mysyr-zegap

World Health Organization

HEALTH EMERGENCIES programme

Confirmation of Participation

SocialNet: Social and behavioural insights COVID-19 data collection tool

noel george

November 29, 2023