Latent Growth Curve Modeling for COVID-19 Cases in Presence of Time Variant Covariate

M. S. Panwar1,C. P. Yadav1, Harendra Singh2, Taghreed M. Jawa3, Neveen Sayed-Ahmed3

1*Department of Statistics, Banaras Hindu University, Varanasi, India-221005.*

2*Department of Mathematics, Post Graduate College, Gazipur, India-233001.*

3*Department of Mathematics and Statistics, College of Science, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia*

# Abstract

For the past two years, the entire world has been fighting against the COVID-19 pandemic. The rapid increase in COVID-19 cases can be attributed to several factors. Recent studies have revealed that changes in environmental temperature are associated with the growth of cases. In this article, we modeled the monthly growth rate of COVID-19 cases per million infected in 126 countries using various growth curves under structural equation modeling. Moreover, the environmental temperature has been introduced as a time-varying covariate to enhance the performance of the models. The parameters of growth curve models have been estimated, and accordingly, the results are discussed for the affected countries from August 2020 to July 2021.

*Keywords:* COVID-19, Environmental Temperature, Latent Growth Curve Model, Multi-Phase Model, Time Variant Covariates

# Introduction

The coronavirus disease 2019 (COVID-19) was first reported in Wuhan city, China. Several individuals in Wuhan’s seafood market were identified with unknown viral pneumonia [[23,](#_bookmark39) [47,](#_bookmark63) [7].](#_bookmark23) In the next few months, the virus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), spread to other cities of China and then worldwide. On 30 January 2020, the Director-General of the World Health Organization (WHO) declared the outbreak of COVID-19 to be a public health

*∗*–

*Email addresses:* [chandrap.yadav4@bhu.ac.in](mailto:chandrap.yadav4@bhu.ac.in) (C. P. Yadav1), [harendra059@gmail.com](mailto:harendra059@gmail.com) (Harendra Singh2), [vir.panwar@gmail.com](mailto:vir.panwar@gmail.com) (M. S. Panwar1), [tmjawa@tu.edu.sa](mailto:tmjawa@tu.edu.sa) (Taghreed M. Jawa3), [nevensayd@yahoo.com](mailto:nevensayd@yahoo.com) (Neveen Sayed-Ahmed3)

*Preprint submitted to Computational Intelligence and Neuroscience December 6, 2021*

emergency of international concern. In March 2020, more than one hundred countries were facing challenges due to this virus, and the infection cases from COVID-19 were identified almost all over the world. Since March 2020 there are now specific vaccines available against SARS-CoV-2. In the absence of specific therapeutic drugs or vaccines, controlling the spread of SARS-CoV-2 was nearly impossible, as the health management system of any country was not sufficient enough to deal with this pandemic [[25,](#_bookmark41) [45,](#_bookmark61) [2,](#_bookmark18) [5].](#_bookmark21)

According to the WHO reports [[1],](#_bookmark17) more than 214 million cases had been reported globally by the end of August 2021, out of which around 4.47 million deaths occurred. Here, it was noticed that the growth rate of infected cases and deaths experienced in different regions were dissimilar. A number of causes can affect the growth rate of cases in a region or country, such as the health management system, government policy, environmental factors. At the initial phase, partial or complete lockdown and quarantine played an important role in controlling the spread of the virus. The work by Bacchetti et al. in [[14]](#_bookmark30) showed that lockdown was highly effective in reducing mortality in more polluted areas at the early stage of the pandemic. Moreover, Marquez et al. [[27]](#_bookmark43) concluded that air pollution results in a higher incidence and mortality from COVID-19. Azuma et al. studied the role of various environmental factors in the transmission of SARS-CoV-2 in indoor spaces [[10].](#_bookmark26)

In the first quarter of 2020, in a study of COVID-19 cases and the related meteorological factors in 122 cities of China, no evidence was found that the case counts of COVID-19 will decline when the weather becomes warmer [[46].](#_bookmark62) On the other hand, an earlier study in the laboratory by Casanova et al. [[16]](#_bookmark32) has verified that SARS-CoV can be inactivated rapidly as temperature increases from 4*oC* to 40*oC*. Even though the data were quite limited for the second quarter of 2020, Mandal and Panwar [[26]](#_bookmark42) and Shao et al. [[39]](#_bookmark55) have suspected that the spread of SARS-CoV-2 may also be affected by the change in temperature. Thereafter, many researchers have established the association between the temperature and COVID-19 cases [[36,](#_bookmark52) [31,](#_bookmark47) [35].](#_bookmark51) For specific geographical regions, the relationship among both in the presence of some other factors was investigated [[29,](#_bookmark45) [19](#_bookmark35), [24,](#_bookmark40) [20].](#_bookmark36)

Modeling of respiratory diseases is always of high priority for researchers. More-

over, the outbreak of COVID-19 presented a new challenge for everyone to deal with this situation. In the last few months, various approaches have been utilized to fit the growth of COVID-19 cases over time. Balli in [[11]](#_bookmark27) has proposed a time series prediction model to obtain the disease curve and predict the pandemic trend using machine learning methods. For this purpose, linear regression, multilayer perception, random forest, and support vector machine learning methods are utilized. Furthermore, the susceptible-infected-recovered (SIR) model is a well-known and widely used method for respiratory diseases. The classic SIR model was updated by incorporating four new factors that are crucial in fitting the data of COVID-19 cases [[18].](#_bookmark34) Several works

have modified the SIR model in the same manner [[3,](#_bookmark19) [22,](#_bookmark38) [38].](#_bookmark54)

Using the generalized logistic and generalized Richards model, Wu et al. [[44]](#_bookmark60) have presented the fitting for COVID-19 cases in China; then, a similar exercise was per- formed for the 33 other countries, which were at a less advanced stage at that time. Moreover, several fractional-order dynamical models for the analysis of the virus spread were proposed [[43](#_bookmark59), [42,](#_bookmark58) [21,](#_bookmark37) [9,](#_bookmark25) [6].](#_bookmark22) Few researchers have attempted the model fitting of the dynamics of COVID-19 cases in the presence of environmental temperature. Shi et al. [[40]](#_bookmark56) have used the modified susceptible-exposed-infectious-recovered (M-SEIR) model by incorporating the temperature factor to simulate the COVID- 19 outbreak dynamics in Wuhan. In other studies, they examined the associations between epidemiological parameters of the dynamics of new cases and temperature using an autoregressive integrated moving average (ARIMA) model [[17].](#_bookmark33) Moreover, Shah et al. [[37]](#_bookmark53) have proposed a compartmental mathematical model for the trans- mission dynamics of the COVID-19 under the Caputo fractional-order derivative. The Hilbert-type inequalities play a major role in mathematics for pattern complex analysis, numerical analysis, and qualitative theory of differential equations and their implementation [[4,](#_bookmark20) [8,](#_bookmark24) [34](#_bookmark50)].

Generally, a time series, cross-sectional or longitudinal data-based approach is utilized when a response variable is observed with respect to time. These methods have suitability concerns and accordingly advantages and disadvantages. In this article, we use structural equation modeling (SEM) with longitudinal data. These models are generally known as latent curve or growth curve models (GCMs). The rest of the article is organized as follows. In Section [2,](#_bookmark0) various facts have been explored using appropriate plots for cases per million (CPM) and temperature over the months. Then, in Section [3,](#_bookmark8) we build various GCMs for all country data and select the most suitable one for further analysis. In Section [4,](#_bookmark13) the temperature has been added as a time-varying covariate in the modeling to enhance the performance of the considered GCM. In Section [5,](#_bookmark14) all results are discussed with their interpretation. Furthermore, the complete article has been summarized and concluded in Section [6.](#_bookmark16)

# Exploratory Data Analysis

In this study, the data for global COVID-19 cases have been obtained from [https:](https://ourworldindata.org/)

[//ourworldindata.org](https://ourworldindata.org/). A total of 126 countries have been considered for cases recorded from August 2020 to July 2021. The CPM given in a month represents the number of cases recorded on the fifteenth day of that month. Accordingly, the monthly temperature is collected for the capitals of all considered countries from [https://www.weather-atlas.com](https://www.weather-atlas.com/). The value representing temperature in a month is the average temperature in that month.

Before starting the analysis, let us explore and discuss some hidden facts about the data. A simple monthly trajectory plot from August 2020 to July 2021 for all countries is given in Figure [1.](#_bookmark1) In this duration, it is quite easy to observe that the growth of CPM in all countries is high in the first month and then stabilizes in most countries in the next few months. Nevertheless, many countries have experienced sudden rapid growth in CPM in the last few months of the year.

150000



100000

Cases per Million

50000

0

Aug,2020 Sep,2020 Oct,2020 Nov,2020 Dec,2020 Jan,2021 Feb,2021 Mar,2021 Apr,2021 May,2021 Jun,2021 Jul,2021

Time(in Months)

Figure 1: Trajectory plot for CPM over the months from August 2020 to July 2021

Figure [2a](#_bookmark2) shows a set of box plots to understand the nature of the data over the

months. The box plots in this figure show the CPM distribution over the months. It can be seen that the median and mean of CPM increase over the months and the mean is significantly larger than the median in all months. Thus, the distribution is positively skewed in all months. Moreover, the median and dispersion increase at a large scale over the months. In a few countries, CPM are very high, so these countries act as outliers in the first few months; however, in the last months, almost all match with the nature of the sample. A correlation matrix plot of CPM over the months is also shown in Figure [2b.](#_bookmark3) Except for the first three months, the correlation is high for months close together in time, but the correlation tends to decrease with increasing the time separation between the measurement months. On the other hand, in the first few months, the correlation decreased for the upcoming months but again

started to increase. This is weak evidence; however, it is the very first indication that seasons may correlate with the growth of CPM. Moreover, a few basic statistics to understand the characteristics of the observed data are given in Table [1.](#_bookmark4)

150000

100000

Cases per Million

50000

0

Aug,2020 Sep,2020 Oct,2020 Nov,2020 Dec,2020 Jan,2021 Feb,2021 Mar,2021 Apr,2021 May,2021 Jun,2021 Jul,2021

Time(in Months)

Aug\_2020

Aug\_2020

Sep\_2020

Oct\_2020

Nov\_2020

Dec\_2020

Jan\_2021

Feb\_2021

Mar\_2021

Apr\_2021

May\_2021

Jun\_2021

Jul\_2021

Sep\_2020

Oct\_2020

Nov\_2020

Dec\_2020

Jan\_2021

Feb\_2021

Mar\_2021

Apr\_2021

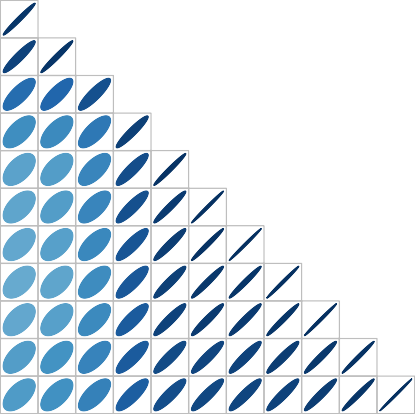
May\_2021

Jun\_2021

Jul\_2021

1

0.8



**1.00 0.98 0.93 0.77 0.61 0.54 0.53 0.52 0.50 0.51 0.55 0.56**

**1.00 0.97 0.80 0.63 0.56 0.56 0.55 0.53 0.55 0.59 0.60**

**1.00 0.87 0.71 0.65 0.65 0.64 0.62 0.63 0.67 0.68**

**1.00 0.94 0.89 0.88 0.86 0.84 0.84 0.83 0.83**

**1.00 0.98 0.96 0.95 0.93 0.92 0.90 0.88**

**1.00 0.99 0.98 0.97 0.95 0.92 0.90**

**1.00 0.99 0.98 0.96 0.92 0.91**

**1.00 0.99 0.98 0.94 0.93**

**1.00 0.99 0.96 0.95**

**1.00 0.99 0.98**

**1.00 1.00**

**1.00**

0.6

0.4

0.2

0

−0.2

−0.4

−0.6

−0.8

−1

* 1. Box plots for CPM over the months
  2. Correlation matrix plot for CPM over the months

Figure 2: A set of box plot and correlation matrix plots of CPM from August 2020 to July 2021 for all countries.

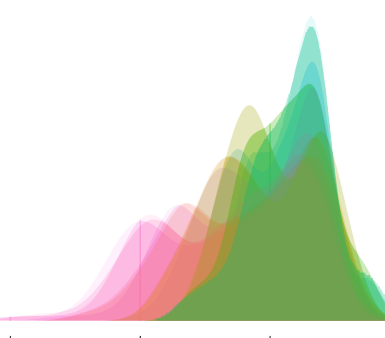
Table 1: Basic statistical measures for CPM over the months.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Aug, 2020 | Sep, 2020 | Oct, 2020 | Nov, 2020 | Dec, 2020 | Jan, 2021 | Feb, 2021 | Mar, 2021 | Apr, 2021 | May, 2021 | June, 2021 | July, 2021 |
| Mean | 3461.19 | 4601.82 | 6300.93 | 10244.54 | 14673.75 | 18986.02 | 22494.01 | 25658.11 | 30217.82 | 34152.76 | 37279.40 | 39993.10 |
| Median | 1336.83 | 2137.14 | 3972.78 | 6961.24 | 9917.47 | 12441.64 | 13814.11 | 15031.02 | 16929.84 | 19220.02 | 24037.40 | 27130.06 |
| Standard Deviation (S.D.). | 4761.05 | 6104.34 | 7820.51 | 11311.03 | 15723.94 | 20345.31 | 23985.04 | 26961.99 | 31248.33 | 34687.31 | 37657.38 | 39044.99 |
| Range | 27279.48 | 36215.94 | 45213.40 | 49765.78 | 67480.17 | 77670.9 | 86453.79 | 96455.48 | 110589.5 | 119737.5 | 152931.6 | 157212.4 |

The distribution of global temperature over the months is given in Figure [3.](#_bookmark5) It can be seen from Figure [3a](#_bookmark6) that the global temperature distribution shifts upwards in the first half of the year and then in the second half it goes downwards. But it does not mean that all the countries follow the same pattern. This can be observed from Figure [3b](#_bookmark7) in which for each month a density plot has been sketched. The density plots are multi-modal because over the months each country contributes to temperature distribution according to its position on the global map. In general, as latitude increases the temperature decreases. This latitude-wise varying pattern

of temperature may have a significant impact on the growth of CPM for respective countries.

40



0.06

20

0.04

Temperature(°C)

Density

0

0.02

Months

Mar,2020 Apr,2020 May,2020 Jun,2020 Jul,2020 Aug,2020 Sep,2020 Oct,2020 Nov,2020 Dec,2020 Jan,2021 Feb,2021 Mar,2021 Apr,2021 May,2021 Jun,2021 Jul,2021

−20

0.00

Mar,2020 Apr,2020 May,2020 Jun,2020 Jul,2020 Aug,2020 Sep,2020 Oct,2020 Nov,2020 Dec,2020 Jan,2021 Feb,2021 Mar,2021 Apr,2021 May,2021 Jun,2021 Jul,2021

Time(in Months)

* + 1. Box Plot for Temperature.

−20 0 20 40

Temperature(°C)

* + 1. Density Plot for Temperature.

Figure 3: A set of box plots and density plots of temperature from August 2020 to July 2021 for all

countries.

# Model Building and Elicitation

The traditional methods for studying the changes in the linear and non-linear framework are regression and analysis of variance (ANOVA). These approaches basically deal with the mean level differences and among individuals, changes are observed from residuals. To utilize the information from residuals several methods like random effect ANOVA, multi-level modeling and hierarchical linear modeling have been proposed. These models explore the differences among individuals with the help of random coefficients. But the limitation of such models is that they are based on a single response variable. A single response variable is not able to capture all the complexities of a growth model [see [[30]].](#_bookmark46)

As the objective of this study is to observe the intraindividual changes and interindividual differences for all countries over the considered time period. So, to possess such characteristics a structural equation modeling for longitudinal data has been proposed. GCMs, which are generally applied for modeling in social and behavioral sciences are used for studying such changes. For GCM, response variables are observed over the ordered time periods whereas some time-invariant or time-variant covariates may also be present. The basics of GCM and hypothesis testing for individual change and interindividual differences are discussed and derived by [[12].](#_bookmark28) In the article [[33],](#_bookmark49) authors have discussed GCM for different models and analyzed the cortisol production data. For more details about longitudinal studies using growth

curve models, one may refer to [[28],](#_bookmark44) [[41],](#_bookmark57) [[15],](#_bookmark31) [[32].](#_bookmark48) For understanding the models based on latent variables with respective R code one may follow [[13].](#_bookmark29)

As the primary objective of this study is to find a better model for CPM trajectory as shown in Figure [1.](#_bookmark1) So, in this section, some possible GCMs such as linear, exponential, latent, and multi-phase have been introduced which could provide a better substitute for CPM fitting. The general structure of the GCM can be given as

*C*[*t*]*n* = Λ0[*t*]*.τ*0*n* + Λ1[*t*]*.τ*1*n* + *· · ·* + Λ*k*[*t*]*.τkn* + *ε*[*t*]*n,* (1) where [*t*]*n* is a multi-occasion vector which represents the observed value of CPM for

*C*

*nth* country for *tth* month. The set of vectors Λ0*,* Λ1*, ,* Λ*k* collectively responsible for the intraindividual change, i.e. each of these captures the growth of CPM in a country over the months. This vector defines the shape of the interindividual change like linear, exponential, etc. for a country. The latent or unobserved variables which are denoted by *τ*0*, τ*1*, , τk*, define interindividual differences in intraindividual change among countries. In the defined model by [(1),](#_bookmark9) each interindividual difference variable *τ*0*, τ*1*, · · · , τk* is associated with the corresponding intraindividual change variable Λ0*,* Λ1*, · · · ,* Λ*k*.

*· · ·*

*· · ·*

Generally, the set of latent variables *τ*0*, τ*1*, · · · , τk* has a multivariate normal distribution with mean vectors (*µ*0*, µ*1*, · · · , µk*) and random variances and covariances *σij*; *i, j* = 1*,* 2*, , k*. The mean vector captures the pattern of intraindividual change and variances and covariances represent the extent to which countries differ within and between. The time-dependent residual variable, *ε*[*t*]*n*, is assumed to have a mean 0 and the same variance, *σ*2 at each occasion, and also it is assumed to be uncorrelated

*ε*

*· · ·*

with other variables.

Let first describe and derive some GCMs using [(1)](#_bookmark9) and then choose the best- fitted model for CPM among them. Few appropriate GCMs which are considered for trajectory fitting are linear, exponential, latent and multi-phase.

* 1. *Linear Growth Curve Model*

In a linear GCM, the growth of the outcome variable is in the form of a straight line which may be in a positive, negative or constant direction over the time periods. A linear GCM can be described by two vectors, Λ0 and Λ1 for different countries over the months from model in [(1).](#_bookmark9) As the model is applied from August 2020 to July 2021, hence we have

Λ0[*t*] = [1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1]

and Λ1[*t*] = [0*,* 1*,* 2*,* 3*,* 4*,* 5*,* 6*,* 7*,* 8*,* 9*,* 10*,* 1]*.*

Here, Λ0 is used to describe the initial level of measurement of the outcome variable when the other effects are 0, whereas Λ1 is responsible for the growth or decline in

*Cn*. This means that the countries can differ from each other in two ways such as their latent intercept (*τ*0) and latent slope (*τ*1). All entries for Λ0 are fixed to 1, this means that intercept affects all measures with equal scores across months.

* 1. *Quadratic Growth Curve Model*

In general the changes over time in measurement variable is non-linear. So, we want to introduce more complexity in the model to capture this non-linearity. For this purpose, we are introducing another vector Λ2, which is responsible for quadratic change in the intraindividual change and interindividual differences. Here, another two vectors Λ0 and Λ1 are responsible for the intercept and linear change in trajectory. We define three vectors as: Λ0[*t*] = [1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1], Λ1[*t*] =

[0*,* 1*,* 2*,* 3*,* 4*,* 5*,* 6*,* 7*,* 8*,* 9*,* 10*,* 1] and Λ2[*t*] = (Λ1[*t*])2 = [0*,* 1*,* 4*,* 9*,* 16*,* 25*,* 36*,* 49*,* 64*,* 81*,* 100*,* 1].

The initial level amount of outcome variable is depicted by *τ*0 and after then at each successive time periods the linear and quadratic changes are governed by vectors *τ*1 and *τ*2.

* 1. *Exponential Growth Curve Model*

In exponential GCM, the two vectors, Λ0 and Λ1, are responsible for the exponential intraindividual change and these can be defined such as

Λ0[*t*] = [1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1]

and Λ1[*t*] = [*e−λ*[0]*, e−λ*[1]*, e−λ*[2]*, e−λ*[3]*, e−λ*[4]*, e−λ*[5]*, e−λ*[6]*, e−λ*[7]*, e−λ*[8]*, e−λ*[9]*, e−λ*[10]*, e−λ*[11]]*,*

where *λ* can be estimated from the observations. The interindividual change among

the countries are depicted by two latent random variables *τ*0 and *τ*1. Here, the random variable *τ*0*n* can be interpreted as the maximum level of *Cn*. The sum of latent slope score and latent asymptotic score, (*τ*0*n* + *τ*1*n*), represents the value of *Cn*[0]. The random variable *τ*1*n* represents a country’s potential for change in *n*[*t*] from initial level to upcoming months. The parameter *λ* indicates the rate at which the level of *Cn*[*t*] changed to the asymptotic level, and here is modeled as being identical for all countries, meaning that the rate at which any individual’s *n*[*t*] level changes is unidirectional (either continuously increasing or decreasing toward his or her asymptotic capacity, *τ*0*n*) and constant (exponentially) across the entire observation period from August 2020 to July 2021. This assumption can be relaxed in further studies.

*C*

*C*

* 1. *Latent Growth Curve Model*

The basis coefficients for a latent GCM are estimated freely so that the optimal change in trajectory can be achieved as per the nature of data. Whereas in earlier discussed models the basis coefficients have been fixed in advance. Here, the basis

coefficients are defined such as

Λ0[*t*] = [1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1]

and Λ1[*t*] = [0*, λ*1*, λ*2*, λ*3*, λ*4*, λ*5*, λ*6*, λ*7*, λ*8*, λ*9*, λ*10*,* 1]*.*

We fixed the first and last basis coefficients as 0 and 1, as it is necessary for the model identification. In the case of latent GCM, the non-linear pattern of intraindividual change is captured by vector Λ1[*t*] and a single interindividual difference variable *τ*1.

* 1. *Multi-Phase Growth Curve Model*

A multi-phase GCM is based on different spline regression models that are connected for different time slots. As it is observed that various countries are facing a number of COVID-19 waves so a multi-phase model may be a good choice for CPM modeling. As Figure [1](#_bookmark1) shows such pattern where the rate of change of Covid-19 cases is not uniform over considered months. From many possible multi-phase GCMs, particularly, *MP*[3*,*4*,*5], has been taken for modeling. The suffix vector specifies the length of phases considered in the model. As the data is taken from August 2020 to July 2021 hence the vector [3,4,5] denotes three phases are taken as Phase-I: (August 2020, September 2020, October 2020), Phase-II: (November 2020, December 2020, January

2021, February 2021), and Phase-III: (March 2021, April 2021, May 2021, June 2021, July 2021). In the *MP*[3*,*4*,*5] model, Phase-I is known as baseline phase and modeled via Λ0 whereas the Phase-II and Phase-III are modeled via Λ1 and Λ2, respectively. Accordingly, the three intraindividual change vectors becomes

Λ0[*t*] = [1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1]*,*

Λ1[*t*] = [0*,* 0*,* 0*, λ*3*, λ*4*, λ*5*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1*,* 1]

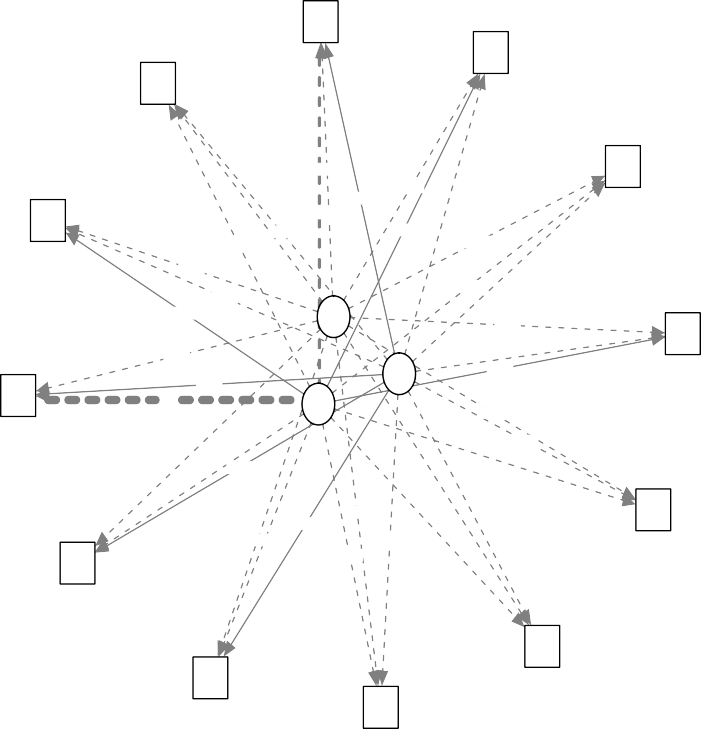
and Λ2[*t*] = [0*,* 0*,* 0*,* 0*,* 0*,* 0*,* 0*, λ*7*, λ*8*, λ*9*, λ*10*,* 1]*,*

where the values of *λjs* can be estimated from the data.

The interindividual difference is governed by the latent random variables *τ*0, *τ*1 and *τ*2. The means of the latent variables *τ*0, *τ*1 and *τ*2 represent the average baseline *C*[*t*]*n* level, amount of [*t*]*n* change in second phase, amount by which [*t*]*n* gains in the last phase, respectively. Simultaneously, the variances of the latent variables represent the extent to which countries differ in these aspects of intraindividual change and how interindividual differences in one aspect are related to interindividual differences in the other aspects can be defined by their covariances.

*C C*

To find the best among the considered models, a few well-known model fitting criteria like Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA) and *χ*2-statistic with degrees of freedom (D.F.), have been taken. Lower values of AIC,



C[10]

C[5]

C[12]

1.00

C[7]

1.00

1.00

1.23

C[4]

1.00

0.00

1.00 0.61

1.00 1.00

1.00 0.00

0.00

1.00

0.27

0

1.00

C[6]

1.00

0.00

2 0.85

C[9]

0.76

1.00

1

1.00

1.00

0.00

0.22

0.00

1.00

1.00

1.00

0.00

1.36

1.00

C[2]

0.00

1.00

0.00

C[8]

0.00

C[1]

C[11]

C[3]

Figure 4: Structure plot for *MP*[3*,*4*,*5].

BIC, RMSEA, *χ*2-statistic and higher values of TLI statistic indicates the choice of a

suitable model. Based on these criteria, it can be seen from Table [2](#_bookmark11) that the multi- phase model, *MP*[3*,*4*,*5] is performing better than others. So, it can be concluded that multi-phase GCM is most appropriate among the considered models. The estimates of the coefficients for all considered GCMs are given in Table [3](#_bookmark12) and discussed in Section [5.](#_bookmark14) For the multi-phase GCM, the structure plot is given in Figure [4.](#_bookmark10) In the structure plot, the dotted lines are showing the fixed factor loading and the dark line is showing the estimated elements of factor loading. It is expected that with the addition of suitable covariates, the overall performance of the multi-phase GCM may also improve.

Table 2: Different fitting criteria for various GCMs of all countries data from August 2020 to July 2021.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | AIC | BIC | TLI | RMSEA | *χ*2-statistic(D.F.) |
| Linear | 18993.40 | 19045.03 | 0.47 | 0.19 | 497.96 (73) |
| Quadratic | 18790.03 | 18853.81 | 0.71 | 0.14 | 286.59 (69) |
| Exponential | 19093.62 | 19148.29 | 0.34 | 0.22 | 596.19 (72) |
| Latent | 18916.38 | 18998.37 | 0.51 | 0.19 | 400.94 (63) |
| Multi-Phase | 18702.93 | 18787.96 | 0.82 | 0.11 | 185.49 (62) |

Table 3: Estimate of coefficients for various GCMs for all countries data from August 2020 to July 2021.

Linear Quadratic Exponential Latent Multi-Phase

*τ*1 *↔ τ*1 22.4033\* 383.8042\* 4.2484 402.1072 2508.8822\* *τ*2 *↔ τ*2 3.4922\* 1428.2694\* *τ*0 *↔ τ*1 -46.4837\* -446.6204\* -18.0949 -268.0263 -605.1945\* *τ*0 *↔ τ*2 39.8306\* 35.6665

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Intercept and Slope: |  | | | | |
| *τ*0 | 77.1223\* | 77.5197\* | 77.5291\* | 77.2063\* | 77.0714\* |
| *τ*1 | -0.0554 | -0.3052 | -0.0486 | -0.2633 | -0.5005\* |
| *τ*2  *λ* (Rate) |  | 0.0240 | 15.9286\* |  | 0.2096\* |
| *λ*1 |  |  |  | -0.1469 |  |
| *λ*2  *λ*3 |  |  |  | 0.3803  0.7973\* | 0.2749\* |
| *λ*4  *λ*5 *λ*6 *λ*7 *λ*8 *λ*9 *λ*10 |  |  |  | 1.6157\*  2.1637\*  2.4558\*  2.4272\*  1.9969\*  1.5036\*  1.1624\* | 0.0705\*  0.8472\*  0.2223\*  0.7587\*  1.2315\*  1.3624\* |
| Covariances: |  |  |  |  |  |
| *τ*0 *↔ τ*0 | 231.1603\* | 634.8185\* | -7.1813 | 322.9.86\* | 354.4765\* |

*τ*1 *↔ τ*2 -35.5733\* -1029.3199\*

\* Significant parameter at p*<*0.05

# Modeling with Time Varying Covariate

Many factors are influencing the growth of CPM but the environmental tempera- ture has its own significant impact. Earlier in Section [2,](#_bookmark0) it has been shown that there may be some dependency between CPM and temperature. Also from Figure [2a](#_bookmark2) and Figure [3a,](#_bookmark6) it can be seen that the monthly gain in CPM depends on the pattern of global temperature. If the temperature is added as a covariate in all country data then the GCMs will perform robustly.

In general, two types of predictor variables are used in longitudinal studies. The covariate which is constant over the measurement time called time-invariant covariate (TIC) and which is varying over time periods called time-varying covariate (TVC). Here, in this study, the temperature is a TVC over the months, which has a direct impact on the CPM with some coefficient, say, *γ*[*t*]. Now the model defined by [(1)](#_bookmark9) can be redefined in presence of TVC as follows

*C*[*t*]*n* = Λ0[*t*]*.τ*0*n* + Λ1[*t*]*.τ*1*n* + *· · ·* + Λ*k*[*t*]*.τkn* + *γ*[*t*]*T* [*t*]*n* + *ε*[*t*]*n.* (2)

Where *T* [*t*]*n* represents the temperature of *nth* country in the *tth* month. All GCMs in presence of TVC are performing better than the respective models introduced in the previous section. Also, we tried different possible combinations of phases to construct

multi-phase models with TVC. Among the considered models, *MPT*

[3*,*4*,*5]

, is outper-

forming with AIC, BIC, TLI, RMSEA and *χ*2-statistic (d.f) values 18722*.*02, 18843*.*50, 0*.*79, 0*.*07 and 361*.*44(194), respectively. In Table [3](#_bookmark12), the estimate of coefficients for all considered GCMs with TVC has been given. For the multi-phase GCM , the structure plot is given in Figure 5. In the structure plot, the dotted lines are showing the fixed factor loadings and the dark lines are representing the estimated elements of factor loadings and estimated coefficients for TVCs.

# Results

In this section, we are discussing about the results from Table [3](#_bookmark12) and Table [4.](#_bookmark15) In Table [3,](#_bookmark12) the estimate of parameters of four GCMs has been given. For linear GCM, the mean baseline level of *C*[*t*]*n* is 77*.*1223 (*τ*0) and then growth is measured with an increment of 0*.*0554 (*τ*1) over the months. In quadratic GCM, the mean baseline level of the outcome variable is 77*.*5197 and 0*.*3052, 0*.*0240 amount of suppressing and increment and in linear and quadratic phases respectively. In exponential GCM, the baseline level is 76*.*9430 which is the sum of *τ*0 and *τ*1. After which growth in [*t*]*n* is observed with an exponential rate of 15*.*9286 to some limit or capacity of a country. The average baseline value from latent GCM is 77*.*2063 and the average total amount of growth is *−*0*.*2633. The estimated value of *C*[*t*]*n* at any month can be calculated by [77*.*2063 + (*−*0*.*2633)*λt*], e.g. in Oct 2020 the estimated value of *C*[*t*]*n* is 76*.*99637. For the considered multi-phase model, *MP*[3*,*4*,*5], the average baseline value is 77*.*0714. The average growth amount from November 2020 to February 2021 is

*−*

*−*

*C*

*−*0*.*5005 and the additional growth amount is 0*.*2096 from March 2021 to July 2021.

Table 4: Estimate of coefficients for various GCMs with TVC for all countries data from August 2020 to July 2021.

Linear Quadratic Exponential Latent Multi-Phase

Intercept and Slope:

*τ*0 75.7481\* 86.0991\* 88.0494\* 82.0764\* 79.8183\*

*τ*1 0.1262 -5.3411 -1.2446\* -2.4983\* -6.8267

*τ*2 0.5547\* 7.8347

*λ* (Rate) 9.8281

*λ*1 -0.1288

*λ*2 0.4034

*λ*3 0.8180\* 0.2743\*

*λ*4 1.6236\* 0.6174\*

*λ*5 2.1601\* 0.0597\*

*λ*6 2.4385\*

*λ*7 2.4080\* 0.2245\*

*λ*8 1.9792\* 0.7487\*

*λ*9 1.4700\* 1.2254\*

*λ*10 1.1387\* 1.3616\*

Covariances:

*τ*0 *↔ τ*0 233.4468\* 625.8157\* -9.0008 313.6485\* 348.3898\* *τ*1*↔ τ*1 22.3028\* 385.5562\* 12.2186\* 411.1636 2485.5481\* *τ*2 *↔τ*2 3.5006\* 1414.3588\* *τ*0*↔ τ*1 -46.4866\* -445.1598\* -34.4772 -272.0185 -596.6951\* *τ*0*↔ τ*2 39.4834\* 21.3108

*τ*1 *↔ τ*2 -35.7064\* -1014.6153\*

Regression (TVC):

*γ*[1] 0.0018 -0.1782 -0.1836 -0.0994 -0.0605

*γ*[2] 0.0316 -0.0524 0.0614 -0.0715 -0.0338

*γ*[3] -0.0002 -0.0289 0.0131 -0.0719 0.0652

*γ*[4] -0.0132 -0.0168 -0.0256 -0.0632 0.0334

*γ*[5] 0.0334 0.0589 0.0363 0.0052 0.0381

*γ*[6] 0.0314 0.0643 0.0376 0.0178 0.0561

*γ*[7] 0.0191 0.0373 0.0251 0.0077 0.0501

*γ*[8] 0.0276 0.0215 0.0371 0.0130 0.0340

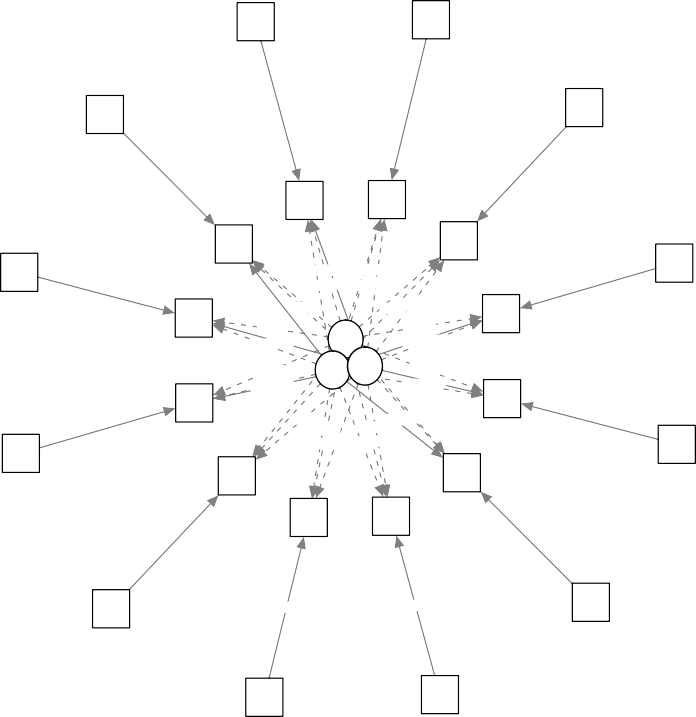
*γ*[9] -0.0005 -0.0346 0.0087 -0.0181 -0.0277

*γ*[10] -0.0228 -0.1161\* -0.0233 -0.0690 -0.1033

*γ*[11] -0.0162 -0.2055\* -0.0138 -0.0757 -0.1268

*γ*[12] 0.0233 -0.0570 0.0256 -0.0318 -0.0539

\* Significant parameter at p*<*0.05



T[8]

T[7]

T[4]

0.03

0.05

T[1]

−0.03

−0.06

C[8]

C[7]

C[4]

C[1]

T[10]

1.00 1.00

T[6]

−0.10

1.00

1.000.221.000.00

1.00

0.06

C[10]

0.207.00

1.00

1.010.23

1.00

1.010.36

1.00

0

1

2

0.00.00

1.00

0.805.00

1.00

1.000.75

C[6]

C[11]

1.00

C[9]

−0.13

1.010.00

1.00

1.000.602.00

−0.03

0.000.000.000.00 T[9]

T[11]

C[12]

C[5]

C[3]

C[2]

−0.05

0.04

T[12]

−0.07

−0.03

T[5]

T[3]

T[2]

Figure 5: Structure plot for *MPT* .

[3*,*4*,*5]

Similar to latent GCM, one can estimate the *C*[*t*]*n* using the estimates of respective

coefficients for the multi-phase model also. The covariances for *τ*0*, τ*1 and *τ*2 for all models are also provided in the table. The variance terms represent the extent to which countries differ at the initial level in intraindividual change and the covariances indicate interindividual differences.

In Table [4,](#_bookmark15) estimates of coefficients are observed for various GCMs in presence of TVC, temperature. In this table, all coefficients can be interpreted in a similar manner as in Table [3,](#_bookmark12) except regression coefficients due to covariate. The regression coefficients can be defined as one unit change in temperature at time *t* is associated with *γ*[*t*] unit change in [*t*]. Here it is noticeable that almost in all months temperature is negatively associated with the growth of CPM.

*C*

# Conclusion

In this article, growth in CPM due to COVID-19 is considered a variable of interest. For different countries, the trajectories of CPM are studied from August 2020 to July 2021. The intraindividual change and interindividual differences were captured using linear, exponential, latent, and multiphase GCMs. Based on certain criteria, the multiphase GCM performs better than the other models. Therefore, it can be preferred for analysis purposes. A number of factors are responsible for the rapid growth of CPM in a country. Moreover, these factors impact different countries with

different weights. Thus, in this study, environmental temperature is considered a co- variate that significantly impacts the growth of CPM. Different GCMs were fitted to the data without and with a covariate. Based on various fitting criteria, it is noticed that GCMs improve when the temperature is introduced as a covariate. So, we can say that temperature may be one of the reasons responsible for the changes in CPM over the months. Nevertheless, other possible factors may have an important role in the growth of CPM and can be included in the model for further study. The inclu- sion of other factors in models may improve results. Furthermore, for the study of growth in CPM for a particular region, there may be differences in the model-based outcomes.

# Acknowledgments

Taif University Researchers Supporting Project number (TURSP-2020/318), Taif University, Taif, Saudi Arabia. This work is partially sponsored by the Institute of Eminence Scheme (IOE/Incentive/2021-22/32443) of the University Grand Commission (UGC), India.

# Declaration of Competing Interest

The authors declare that there is no conflict of interest.

# References

1. World Health Organization: WHO Coronavirus Disease(COVID-19) Dashboard.

URL: [https://covid19.who.int](https://covid19.who.int/).

1. Abd-Elhafiez WM, Amin HH. The digital transformation effects in distance education in light of the epidemics (covid-19) in egypt. Information Sciences Letters 2021;10:141 –52.
2. Abdullah , Ahmad S, Owyed S, Abdel-Aty AH, Mahmoud EE, Shah K, Alrabaiah

H. Mathematical analysis of covid-19 via new mathematical model. Chaos, Solitons & Fractals 2021;143:110585.

1. Ahmed A, AlNemer G, Zakarya M, Rezk H. Some dynamic inequalities of hilbert’s type. Journal of Function Spaces 2020;2020.
2. Al-Ajlan RA, Khalifa WAA, Atta HAY. The role of adopting blackboard system in light of the coronavirus pandemic in developing the skills of its use and aca- demic achievement among taif university students. Information Sciences Letters 2021;10:177 –88.
3. Ali A, Khan MY, Sinan M, Allehiany F, Mahmoud EE, Abdel-Aty AH, Ali

G. Theoretical and numerical analysis of novel covid-19 via fractional order mathematical model. Results in Physics 2021;20:103676.

1. Alnaser WE, Abdel-Aty M, Al-Ubaydli O. Mathematical prospective of coron- avirus infections in bahrain, saudi arabia and egypt. Information Sciences Letters 2020;9:51 – 64.
2. AlNemer G, Zakarya M, El-Hamid A, Hoda A, Agarwal P, Rezk HM. Some dynamic hilbert-type inequalities on time scales. Symmetry 2020;12(9):1410.
3. Ameen IG, Ali HM, Alharthi M, Abdel-Aty AH, Elshehabey HM. Investigation of the dynamics of covid-19 with a fractional mathematical model: A comparative study with actual data. Results in Physics 2021;23:103976.
4. Azuma K, Yanagi U, Kagi N, Kim H, Ogata M, Hayashi M. Environmental factors involved in SARS-CoV-2 transmission: effect and role of indoor envi- ronmental quality in the strategy for covid-19 infection control. Environmental health and preventive medicine 2020;25(1):1–16.
5. Ballı S. Data analysis of Covid-19 pandemic and short-term cumulative case fore- casting using machine learning time series methods. Chaos, Solitons & Fractals 2020;142:110512.
6. Baltes PB, Nesselroade JR. History and rationale of longitudinal research. 1979. [13]Beaujean AA. Latent variable modeling using R: A step-by-step guide. Rout-

ledge, 2014.

1. Becchetti L, Conzo G, Conzo P, Salustri F. Understanding the heterogeneity of adverse covid-19 outcomes: the role of poor quality of air and lockdown decisions. Available at SSRN 3572548 2020;.
2. Bollen KA, Curran PJ. Latent curve models: A structural equation perspective. volume 467. John Wiley & Sons, 2006.
3. Casanova LM, Jeon S, Rutala WA, Weber DJ, Sobsey MD. Effects of air tem- perature and relative humidity on coronavirus survival on surfaces. Applied and environmental microbiology 2010;76(9):2712–7.
4. Demongeot J, Flet-Berliac Y, Seligmann H. Temperature decreases spread pa- rameters of the new Covid-19 case dynamics. Biology 2020;9(5):94.
5. Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, Colaneri M. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nature Medicine 2020;:1–6.
6. Han Z, Li S, Liu H. Composite learning sliding mode synchronization of chaotic fractional-order neural networks. Journal of Advanced Research 2020;25:87–96.
7. Haque SE, Rahman M. Association between temperature, humidity, and COVID- 19 outbreaks in Bangladesh. Environmental science & policy 2020;114:253–5.
8. Khan H, Ibrahim M, Abdel-Aty AH, Khashan MM, Khan FA, Khan A. A frac- tional order covid-19 epidemic model with mittag-leffler kernel. Chaos, Solitons & Fractals 2021;148:111030.
9. Khan Z, Van Bussel F, Hussain F. A predictive model for Covid-19 spread–with application to eight us states and how to end the pandemic. Epidemiology & Infection 2020;148.
10. Kumar A, Singh R, Kaur J, Pandey S, Sharma V, Thakur L, Sati S, Mani S, Asthana S, Sharma TK, Chaudhuri S, Bhattacharyya S, Kumar N. Wuhan to world: The covid-19 pandemic. Frontiers in Cellular and Infection Microbiology 2021;11:242.
11. Liu H, Li SG, Sun YG, Wang HX. Prescribed performance synchronization for fractional-order chaotic systems 2015;24(9):090505.
12. Liu YC, Kuo RL, Shih SR. Covid-19: The first documented coronavirus pan- demic in history. Biomedical Journal 2020;43(4):328–33.
13. Mandal CC, Panwar M. Can the summer temperature drop COVID-19 cases?

Public Health 2020;.

1. Marqu`es M, Rovira J, Nadal M, Domingo JL. Effects of air pollution on the potential transmission and mortality of covid-19: A preliminary case-study in tarragona province (catalonia, spain). Environmental Research 2020;192:110315.
2. McArdle JJ, Epstein D. Latent growth curves within developmental structural equation models. Child development 1987;:110–33.
3. Meraj G, Farooq M, Singh SK, Romshoo SA, Nathawat M, Kanga S, et al. Coronavirus pandemic versus temperature in the context of Indian subcontinent: a preliminary statistical analysis. Environment, Development and Sustainability 2020;:1–11.
4. MUTH E, BO , CURRAN P. General longitudinal modelling of individual dif- ferences in experimental designs: a latent variable framework for analysis and power estimation. Psychological Methods 1997;2:371–402.
5. Pirouz B, Golmohammadi A, Masouleh H, Violini G, Pirou B. Relationship be- tween average daily temperature and average cumulative daily rate of confirmed cases of covid-19. medrxiv. 2020. Google Scholar ;.
6. Qureshi I. Latent growth modeling and latent class analysis. PACIS 2009 Pro- ceedings 2009;:39.
7. Ram N, Grimm K. Using simple and complex growth models to articulate de- velopmental change: Matching theory to method. International Journal of Be- havioral Development 2007;31(4):303–16.
8. Rezk H, AlNemer G, Abd El-Hamid H, Abdel-Aty AH, Nisar KS, Zakarya M. Hilbert-type inequalities for time scale nabla calculus. Advances in Difference Equations 2020;2020(1):1–21.
9. Riddell S, Goldie S, Hill A, Eagles D, Drew TW. The effect of temperature on persistence of SARS-CoV-2 on common surfaces. Virology journal 2020;17(1):1– 7.
10. Rouen A, Adda J, Roy O, Rogers E, L´evy P. COVID-19: relationship between atmospheric temperature and daily new cases growth rate. Epidemiology & Infection 2020;148.
11. Shah K, Khan ZA, Ali A, Amin R, Khan H, Khan A. Haar wavelet colloca- tion approach for the solution of fractional order covid-19 model using caputo derivative. Alexandria Engineering Journal 2020;59(5):3221–31.
12. Shahzad M, Abdel-Aty AH, Attia RA, Khoshnaw SH, Aldila D, Ali M, Sultan F. Dynamics models for identifying the key transmission parameters of the covid-19 disease. Alexandria Engineering Journal 2021;60(1):757–65.
13. Shao W, Xie J, Zhu Y. Mediation by human mobility of the association be- tween temperature and covid-19 transmission rate. Environmental Research 2020;:110608.
14. Shi P, Dong Y, Yan H, Zhao C, Li X, Liu W, He M, Tang S, Xi S. Impact of temperature on the dynamics of the COVID-19 outbreak in China. Science of The Total Environment 2020;:138890.
15. Singer JD, Willett JB, Willett JB, et al. Applied longitudinal data analysis: Modeling change and event occurrence. Oxford university press, 2003.
16. Singh H, Srivastava H, Hammouch Z, Nisar KS. Numerical simulation and sta- bility analysis for the fractional-order dynamics of COVID-19. Results in Physics 2020;:103722.
17. Veeresha P, Gao W, Prakasha DG, Malagi NS, Ilhan E, Baskonus HM. New dynamical behaviour of the coronavirus (2019-ncov) infection system with non- local operator from reservoirs to people. Information Sciences Letters 2021;10:205

–12.

1. Wu K, Darcet D, Wang Q, Sornette D. Generalized logistic growth modeling of the COVID-19 outbreak in 29 provinces in China and in the rest of the world. arXiv preprint arXiv:200305681 2020;.
2. Wu YC, Chen CS, Chan YJ. The outbreak of covid-19: An overview. Journal of the Chinese Medical Association 2020;83(3).
3. Xie J, Zhu Y. Association between ambient temperature and COVID-19 infection in 122 cities from China. Science of the Total Environment 2020;724:138201.
4. Zhu H, Wei L, Niu P. The novel coronavirus outbreak in wuhan, china. Global Health Research and Policy 2020;5(1):6.