

## RESEARCH ARTICLE

## EXPLORING COVID-19 PANDEMIC INITIAL 2020 CURVE BASED ON STATISTICAL EVALUATION

Ibrar ul Hassan Akhtar<sup>a,b</sup><sup>a</sup> Department of Meteorology, COMSATS University Islamabad, Park Road, Tarlai Kalan, Islamabad 45550, Pakistan.<sup>b</sup> Space Applications and Research Wing, Pakistan Space and Upper Atmosphere Research Commission, Islamabad, Pakistan.\*Corresponding Author's email: [ibrar.phd.cit@gmail.com](mailto:ibrar.phd.cit@gmail.com)

This is an open access journal distributed under the Creative Commons Attribution License CC BY 4.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited

## ARTICLE DETAILS

## Article History:

Received 22 December 2022  
Revised 12 January 2023  
Accepted 23 February 2023  
Available online 27 February 2023

## ABSTRACT

This research was to explore CoVID-19 at initial stage of 2020 based on statistical techniques using probability density function (PDF), with different statistical measures, polynomial data fitting along with 30 day projections. The Covid-19 analysis was carried out only for highly affected countries along with six different regions and global level for 64 days covering period of January to March during 2020. It was found that infection and recovery rate for cases were ranged from 0 – 9.89 and 0 – 8.89% at global level, respectively. The PDF was observed highly positive skewed, leptokurtic, for confirmed cases representing 6620 daily mean infected population for confirmed cases. Countries of USA, Chinas, Spain, France, Italy, Iran, UK and Switzerland were expected to be most affected countries with minimum 0.100 million infected population. The projection errors for infection rate remained -78.8 to 49.0%. The curve of CoVID-19 and PDF (skewness and kurtosis) measures provided understanding of data shape and peak height.

## KEYWORDS

COVID-19; Time Series Analysis, Statistical Inferences, Public Health, Modeling

## 1. INTRODUCTION

Zoonotic coronaviruses are grouped into four genera of alpha coronavirus, beta coronavirus, gamma coronavirus and delta coronavirus based on difference on polyproteins, structural and accessory proteins (Cui et al., 2019). The severe acute respiratory syndrome (SAR) coronavirus belongs to Beta coronavirus genus and has capability to shift from natural host of bat, mouse to transitional host like camels, cows, cats and pigs) based on genetic multiplicity and ordinary co-evolution (Cui et al., 2019; Wu et al., 2020). The current coronavirus pandemic (SARs-CoVID-2) mentioned as CoVID-19, not first time for the world, threaten socio-economic denial and hindrance to UN sustainable development goals (SDGs). The H1N1 influenza pandemic 1918 provided an understanding linked to lethal past that had killed approx. 50 to 100 million people (Moren and Fauci, 2007). There is always need to quest question such as (1) origin of virus (2) pathogenesis nature (3) death and age relevancy (4) disease multi-annual pandemic waves (5) predictability of disease and vector cycle? with future scenarios to mitigate such pandemics amongst developed and developing nations (Moren and Fauci, 2007).

There are definite features to be measured for a disease in relative to environment and to distinguish among actual relationship or causation through on its strength, regularity, specificity, temporality, biological slope, likelihood, rationality, experiment and similarity (Hill, 1965). The mitigation measures for a pandemic include (i) an operative use of accessible data (ii) rational valuation for mitigation actions and (iii) social, economic and political calculations such as segregation of sick persons, self-isolations, alteration in daily life practices, restrictions of travel, ban on social congregations, educational institutions closings, masks use and personal shielding equipment's, social distancing and vice versa (Inglesby et al., 2006). No doubt, history of epidemic provides substantial advice, if persons know past and react with knowledge (Jones, 2020). Data gathering, distribution and plotting provide near real time method to

comprehend the temporal, spatial configurations and to support decision and policy makers, health care professional and social community like web link console (Dong et al., 2019).

An operative response to pandemic mandate an integrated method based on medical science, data science, social science, artificial intelligence, statistics, information technology, meteorology, biomedical science, biotechnology, public health, anthropology, logistics, diplomacy, crisis management and so on (Bedford et al., 2019). The Lancet emphasized that data candidness, correctness, consistency and sharing is considered paramount for better reaction towards any health emergency similar to CoVID-19 (Lancet, 2020). Communicable disease similar to CoVID-19 highlight a peculiar requirement to develop operative control strategy through epidemic trends and related impacts on public (Giordano et al., 2020). According to Casella, daily CoVID-19 plague statistics was based on draconian policies from control philosophy for China, UK and Italy (Casella, 2020). The CoVID-19 disease life-threatening physical characteristics embraces fast uneven dynamics, quantity and actuation interruption, ambiguity due to deficiency of information and virus performance, absence of effective vaccines and deprived health co-ordination response.

Time sequence investigation of MERS-CoV grounded on seasonality, trends, and statistical predictions confirmed the use of fundamental dynamics by the policy makers and experts from healthcare sector (Da'ar and Ahmed 2018). The Deb applied autoregressive integrated moving average (ARIMA) technique to discover CoVID-19 pattern of disease incidence, reproduction number and provide explain existing outbreak (Deb, 2020). The disease examination models of an improved suspected, infected and recovered model (SIR) of asymptomatic CoVID-19 outperformed standard SIR model by providing more details for northern Italy (Gaeta, 2020; Smith and Moore, 2004; Sattenspiel, 1987). A Susceptible, Infected, Diagnosed, Healed, Ailing, Recognised, Threatened

## Quick Response Code



## Access this article online

## Website:

[www.actascientificamalaysia.com](http://www.actascientificamalaysia.com)

## DOI:

[10.26480/asm.01.2023.08.16](https://doi.org/10.26480/asm.01.2023.08.16)

and Extinct model called SIDHARTHE also exhibited capability to distinguish patients among identified cases and non-diagnosed ones from Italy. Integrated use of mobile based surveys and AI helped in quick identification of cases that will decrease disease spread amongst high danger population (Rao and Vazquez, 2020).

Although, the growth models built on Weibull and Hill equation, logistics provided understanding of CoVID-19 epidemiology and its statistical nature (Hembran and Hembram, 2021). Two additional models i.e., log-linear on percent change, and linear on unit change are data driven method which helped to comprehend the Hong Kong CoVID-19 and associated gender (Zhao et al., 2020). Another remarkable model which has been established is based on four Chinese cities modelled flight data (passenger's destinations) from FLIRT database to conduct CoVID-19 global risks assessment (Haider et al., 2020). While, some other researches focused on various environmental parameters like temperature, wind speed, relative humidity to highlight fundamental associations and statistically linked (Pearson correlations) with disease spread and establishes temperature to be insignificant to moderately related (Bhattacharjee, 2020). The isolation of cases and contact tracing during outbreak can only be practicable for dropping the extent of pandemic, and manage over lengthier duration through model of branching process (Hellewel et al., 2020).

Recent literature further elaborates the need and approaches which were used to understand, evaluate, simulate and estimate the CoVID-19 across the globe at various spatio-temporal scales. Estimation of the outbreak is considered as one of prime element to minimize the spread through monitoring, tracking and projections (Zhong et al., 2020). SimCoVID open-source simulation computer program based on various models represented by a set of differential-algebraic equations is recently proposed (Abdulrahman, 2023). Program used various statistical methods such as linear, mean, root mean square errors and standard deviations. Another research investigate CoVID-19 based on various statistical processes of descriptive analysis, control charts, Pareto charts and data modeling (Eissa et al., 2022). This approach facilitates quantitative assessment based on simple inexpensive statistical programs by addressing issues of data size, data complexity, interpretation and examination (Cox et al., 2018).

CoVID-19 data needs to be checked for data variance and low counts along with skewness based transformations if required (Jayaraj et al., 2023). The evolution of CoVID-19 can be predicted based on mathematical modeling under the need of analyzing its transmission dynamics (Lee et al., 2022). According to Nadim predictions of CoVID-19 transmission in UK greatly help in capturing trend as well as reduction in new cases (Nadim, 2021). Data analytics tools can also be good way to respond to an epidemic by using exploratory data analysis and inferential methods (Ospina et al., 2022). A group researchers carried out first study on CoVID-19 based on SIR model and claimed that 14 days assumption is not good (Taimoor et al., 2022). SIR model is simple epidemiological model proposed by Kermack and McKendrick in 1927 as is considered as simple deterministic model for future predictions (Anderson, 1991; Wang et al., 2020). Further, Pakistani government strategy minimized the disease burden and avoiding its conversion into a pandemic.

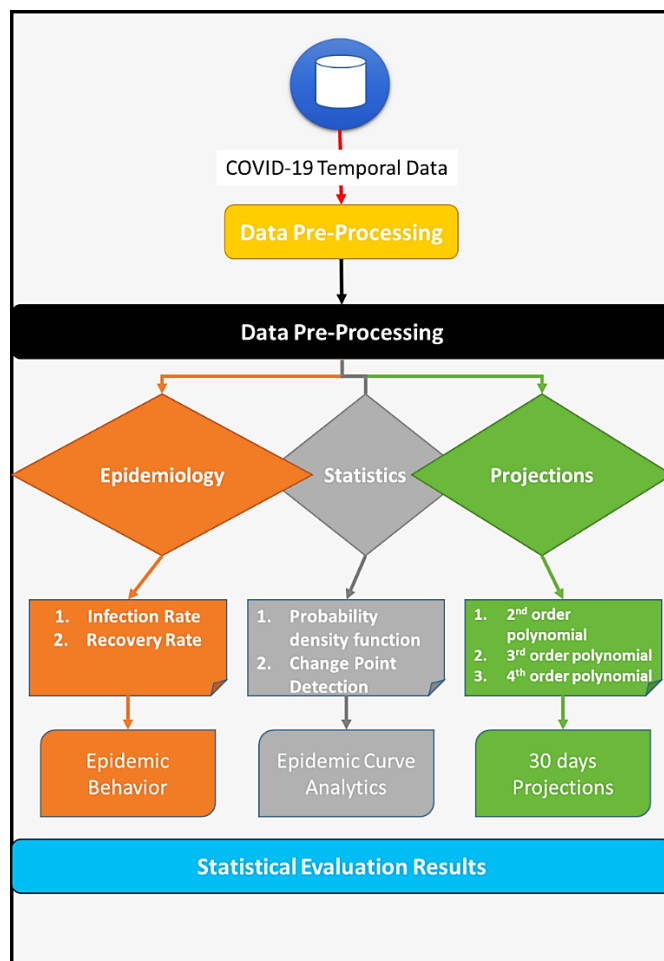
Statistical investigation to study virus infection based on analysis of epidemic data, simulation and option for extrapolation help for better control on CoVID-19 in Italy (Tosto, 2022). Time series data of new, recovered and death cases are used to carry out classical analytical and numerical models based on statistical data (Giuliani et al., 2020; Koo et al., 2020). Prediction models are considered as innovation tools that solve the problem of short term epidemic trend analysis through space and time integrations (Wang et al., 2023). Two techniques of auto-regressive integrated moving average (ARIMA) and Prophet Facebook machine learning are first time used for the CoVID-19 outbreak analysis in Kingdom of Saudia Arabia (Zrieq et al., 2022).

This research was an effort to comprehend the pandemic of CoVID-19 through statistical analytical method based on daily disease cases rates, PDF's, associated dispersions measures, change point events and data fitting curve with polynomial equations along with 30 days projections on regional, global, and above average affected global countries excluding Pakistan. The objectives of this study were as follows: (1) to provide statistical insight of CoVID-19 epidemic (2) to provide short term projections across the globe and (3) to identify the change points events in data for spread of CoVID along with stages in days.

## 2. METHODOLOGY

Statistical modular approach is developed to explore the CoVID-19

pandemic curve evaluation at initial stages during 2020 (Figure 1). Approach include the data pre-processing, epidemiological analysis, statistics evaluation and temporal projections to document the epidemic behaviour, curve analytics and projections. The daily WHO situation reports from January to March 2022 which were collected and data were compiled based on reported cases of confirmed and recovered patients (WHO, 2020). Daily cases data was extracted from the cumulative statistics of WHO data by simple arithmetic of subtracting today from the yesterday.



**Figure 1:** Statistical Modular approach used for CoVID-19 pandemic curve evaluations

Compiled data were used to estimate the infection rate (CIR), recovery rate (CRR), detection of change points in data (CPD), Probability density functions (PDF), polynomial data curve fitting for 2nd to 4th order with projections of thirty days till April 24, 2020 and assessment of polynomial data fitting through difference between projected and observed CoVID-19 cases in term of percentage for April 01, 2020.

$$CIR = \frac{\text{Number of Daily Confirmed Cases}}{\text{Total Confirmed Cases}} \times 100 \quad (1)$$

$$CRR = \frac{\text{Number of Daily Recovered Cases}}{\text{Total Recovered Cases}} \times 100 \quad (2)$$

The Probability density function i.e., PDF was computed based on kernel smoothing procedure and dispersion measures were also estimated for data kurtosis and skewness value based on cases of CoVID-19 confirmed and recovered ones (Chissom, 1970; Wand and Jones, 1995; Duong 2007; Doane and Lori, 2011; Wilks, 2011; Akhtar and Athar, 2019; Akhtar and Athar, 2020; Akhtar, 2020; Akhtar and Athar, 2021).

$$PDF = ks \text{ density } (y, v_j, k) = \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{2\pi}} * \frac{1}{k} * e^{-\frac{(y-v_j)^2}{2k^2}} \quad (3)$$

Where  $y$  was computed for density value,  $v_j$  was kernel median based on the distributed sample,  $k$  ( $> 0$ ) was used as measure for the kernel, and probability density estimate is for PDF computations. To comprehend the daily CoVID-19 data time series, multiple CPD was computed by using binary segmentation technique based on Eq. (iv) (Edwards and Cavalli-Sforza, 1965; Sen and Srivastava, 1975; Killick and Eckley, 2014). There are two available methods for CPD analysis i.e., (i) likelihood ratio for

mean change in data and accumulative data sum based test statistics and (ii) normal distribution based change in data mean and variance (Page, 1954; Hinkley, 1970; Picard et al., 2005).

$$\sum_{i=1}^{m+1} [C(y(\tau_{i-1}+1):\tau_i)] + \beta f(m) \tag{4}$$

where  $C$  is a cost function for a segment e.g., negative log-likelihood and  $\beta f(m)$  is a penalty to guard against over fitting (a multiple change point version of the threshold  $c$ ). A brute force approach to solve this minimization considers  $2^{n-1}$  solutions reducing to  $\binom{n-1}{m}$  if  $m$  is known.

The polynomial equations of second to fourth order were fitted to the WHO data based on best fit to data points around the curve. This helped to comprehend behaviour of data points around the curve for WHO reported CoVID-19 cases with future predictions. The polynomial equations were computed by using succeeding calculations (Perperoglou et al., 2019);

$$y_i = \alpha_0 + \alpha_1 d_i + \alpha_2 d_i^2 + \epsilon \tag{5}$$

$$y_i = \alpha_0 + \alpha_1 d_i + \alpha_2 d_i^2 + \alpha_3 d_i^3 + \epsilon \tag{6}$$

$$y_i = \alpha_0 + \alpha_1 d_i + \alpha_2 d_i^2 + \alpha_3 d_i^3 + \alpha_4 d_i^4 + \epsilon \tag{7}$$

Where  $y_i$  is fitting curve function,  $\alpha_0$  represent intercept,  $\alpha_1$  to  $\alpha_4$  are slope coefficients and  $d_i$  is function of  $x$  called basis function which is defined as:

$$d = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & x_1^4 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 & x_n^4 \end{bmatrix} \tag{8}$$

### 3. RESULTS AND DISCUSSION

The CoVID-19 research study results are documented in this chapter which include the epidemic behaviour, Curve analytics and Polynomial regressions and projections. This will highlight the initial pandemic

nature, extent of impacts on global health system and possible way to minimize its impacts on local, regional as well as global communities.

#### 3.1 COVID-19 Epidemic Behaviour

The global CoVID-19 CIR ranged between 0 to 3.63 percent with dichotomous in nature during January to February, 2022 which later surge to 9.89 percent on March 23, 2020 (Figure 1). The region of western pacific showed dominating characteristics in global CIR by contributing up to 100 percent until February 20th, 2020. While European seemed to be second hotspot area with contribution of 28 percent to global daily CIR on February 23rd, 2020; with CIR maximum of 84.04 percent on March 13<sup>th</sup>, 2020. The region of Eastern Mediterranean was detected as 3rd region impacted by CoVID-19 with 33.88 percent CIR on March 3<sup>rd</sup>, 2020; CIR maximum of 45.47 percent on March 12th, 2020. The American observed as fourth hotspot region with 17.36 percent CIR on March 12th, 2020; with CIR maximum of 24.36 percent on March 23rd, 2020. The regions of South East Asia and African persisted as least affected due to CoVID-19 pandemic with below 5 percent CIR till March, 2020. The CIR of Western Pacific region started decreasing suddenly on February 20th and reached under 5 percent on March 13th, 2020.

Nevertheless, global CRR indicated delay with stretched curves seemed around 1st week of February and touched maximum 8.89 percent on March 24th, 2020; due to constant influence by Western Pacific region. Global countries except Pakistan with above average CoVID-19 impacted across 06 regions CIR and CRR data were evaluated (Figure 2 and Figure 3). The results showed that China was highest in term of infected population on average with 55.2 percent at global level followed by 12.0 percent in Italy, 9.0 percent in South Korea, 7.0 percent in Iran, 4.8 percent in USA, 4.7 percent in Spain, 4.0 percent in Germany, 2.6 percent in France and remaining countries with under 2.0 percent. Republic of China showed sharp recovery rate for CoVID-19 with 88.0 percent CIR value followed by Iran and Italy with 5 percent and remaining countries observed with under 2 percent; due to late onset of native transmission. Pakistan displayed CIR of under 0.1 percent owing to appearance of first case on February 27th, 2020 and closed to zero CRR due to recovery cases happening from March 8th, 2020.

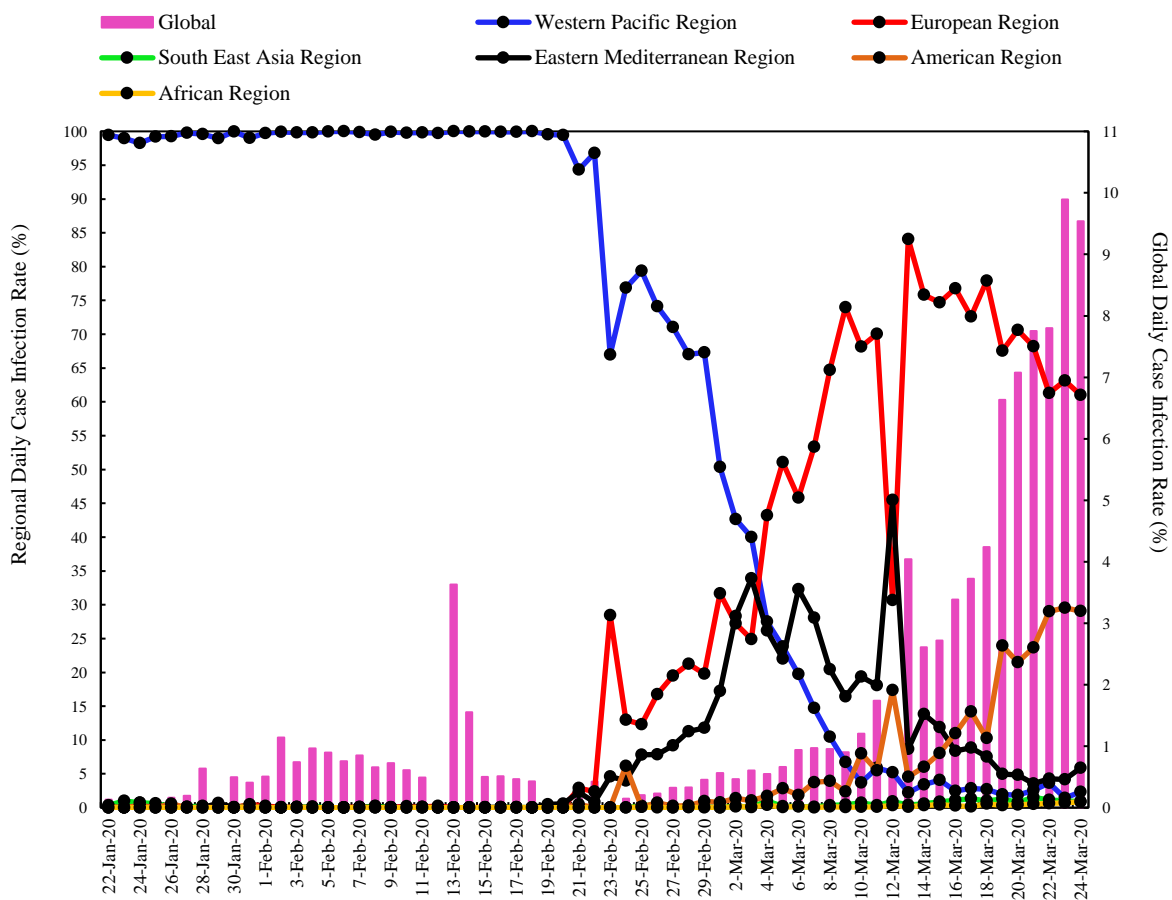


Figure 2(a): The CoVID-19 Global and Regional Corona Infection Rate.

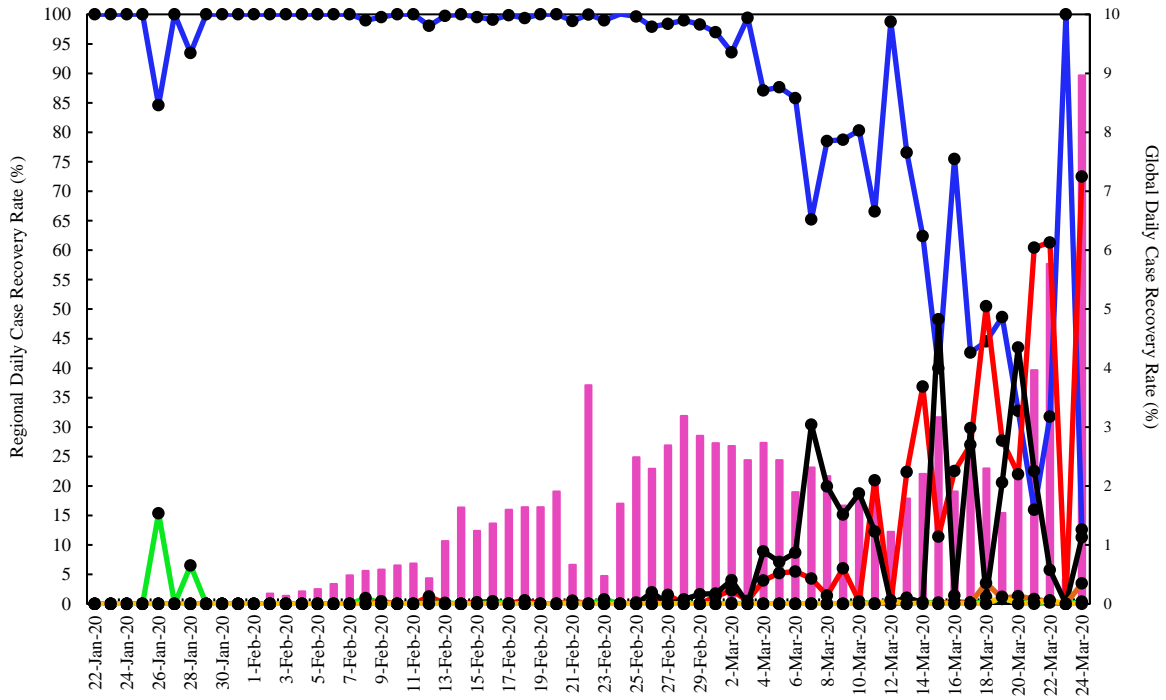


Figure 2(b): The COVID-19 Global and Regional Corona Infection Rate.

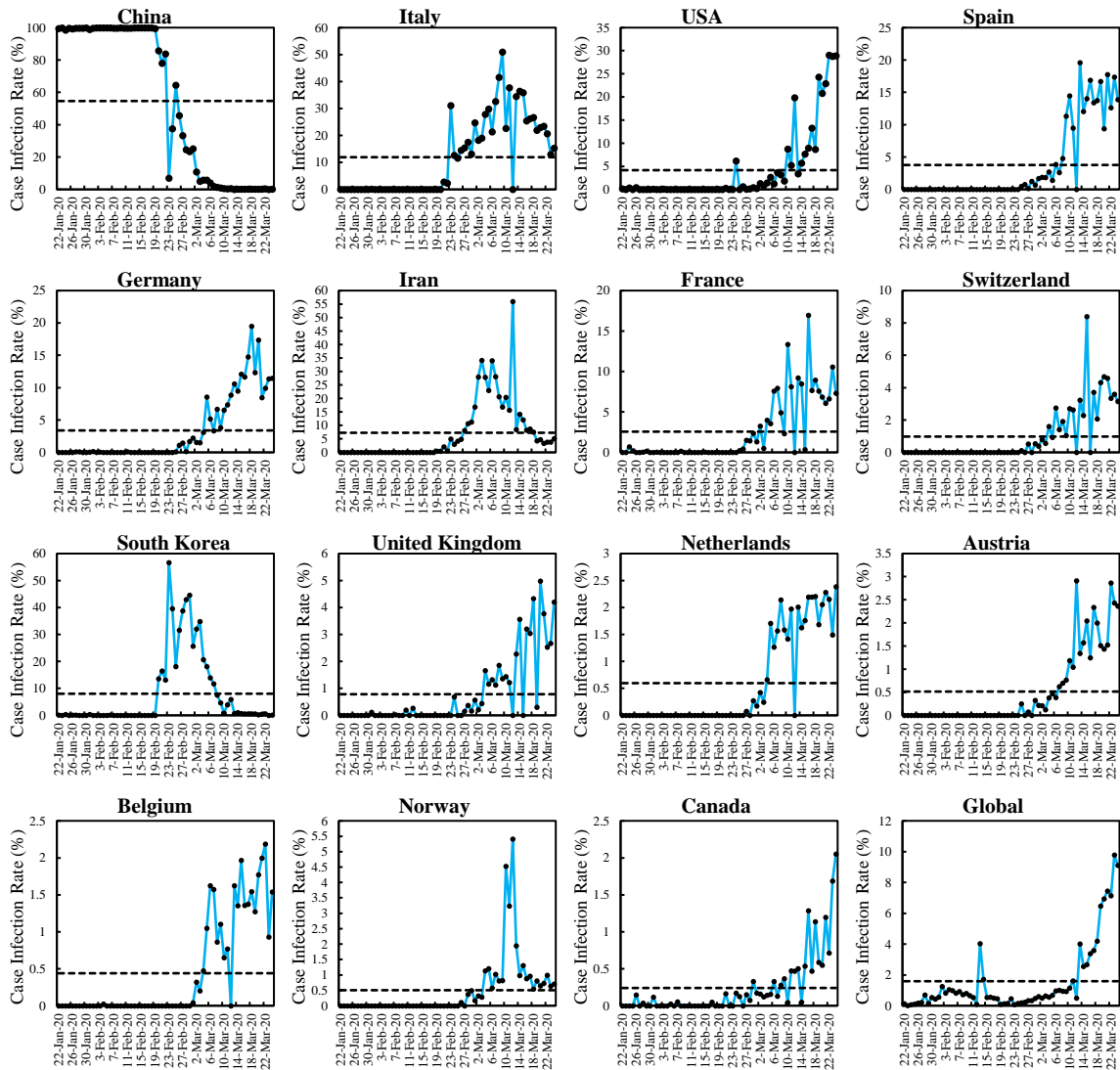


Figure 3: The COVID-19 Infection Rate across global above average affected countries.



### 3.2 Epidemic Curve Analytics

CoVID-19 confirmed cases time series data were evaluated grounded on statistical method to understand the pandemic curve. The daily cases of confirmed and recovered patients PDF's and different statistical measures (values of skewness and kurtosis) showed distinguishing patterns at regional scale (Appendix Figure 2 and 3). Confirmed cases of all regions confirmed were highly positive skewed with values of 1.6 - 4.3) except for Eastern Mediterranean region with bimodal approximate symmetrical curve with value of 0.4). Western Pacific region showed kurtosis of 26.6

with mean of 1541, leptokurtic high and sharp peaks in nature, which was followed by American with kurtosis value of 6.7 and mean = 1011, South East Asia with 5.2 and mean of 50, African with value of 4.9 and mean of 25 and European region with 4.3 and mean of 3539. However, Platykurtic lower curve with wider peak was detected for region of Eastern Mediterranean with kurtosis value of 6.7 and mean of 454. Overall confirmed cases presented high positive skewed with value of 2.2 and leptokurtic with value of 6.8 and mean of 6620 with peak of extreme values over tail right side at global level.

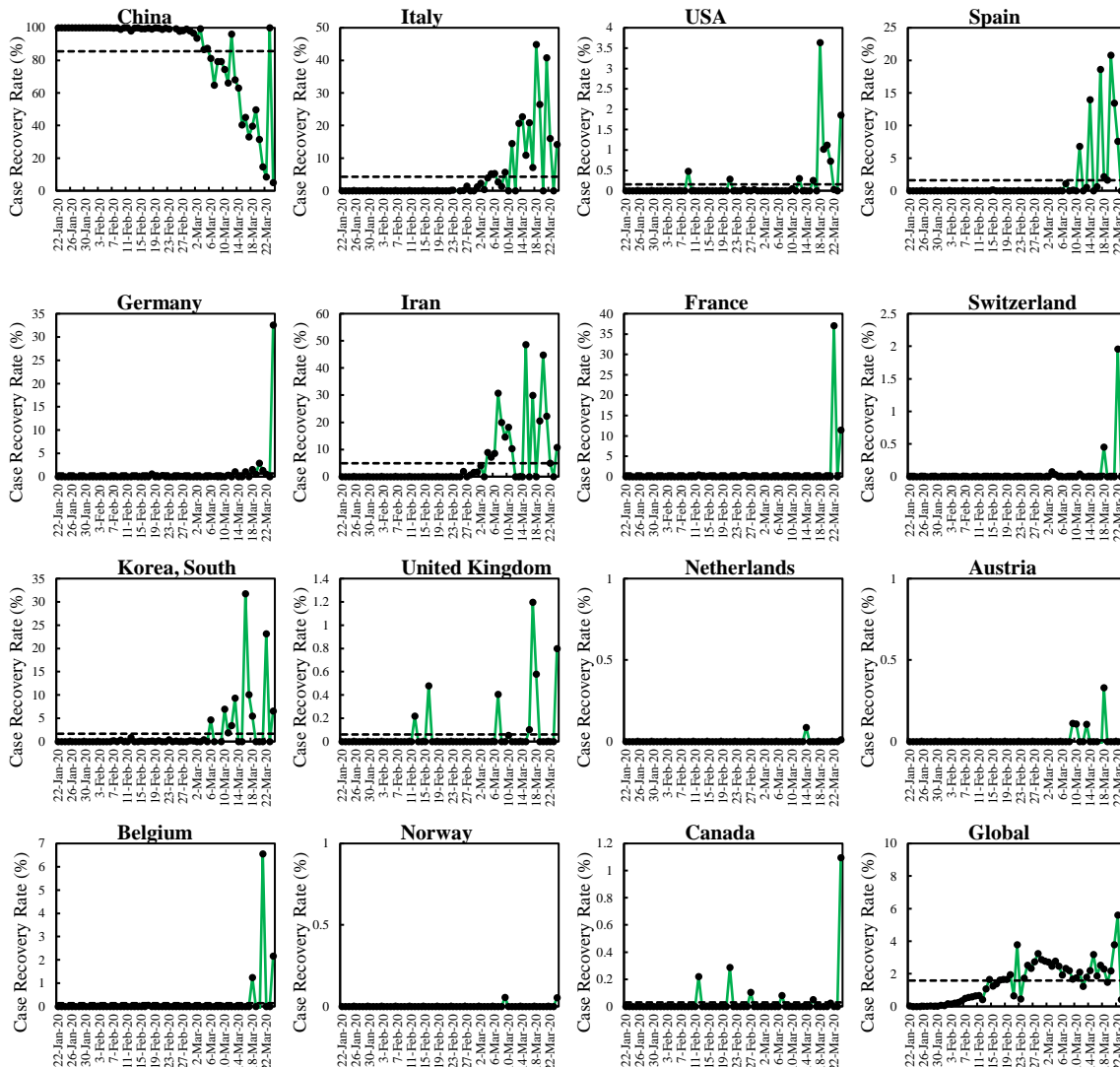


Figure 4: The CoVID-19 Recovery Rate across global above average affected countries.

The regional recovered cases were high positive skewed with skewness value of 1.5 - 3.7) except for Western Pacific with moderately bimodal skewed with skewness value of 0.6. European region showed high leptokurtic and high-pitched peaks with kurtosis value of 17.1 and mean of 315), followed by region of South East Asian with kurtosis value of 15.0 and mean of 2, American region with kurtosis value of 14.6 and mean of 9, African region with kurtosis value of 4.9 and mean of 1) and Eastern Mediterranean region with kurtosis value of 4.6 and mean of 150. Estimated Mesokurtic peak was detected for region of Western Pacific with kurtosis value of 2.7 and mean of 1231. The recovered cases observed very bimodal high positive skewed with value of 2.0 and leptokurtic with kurtosis value of 10.0 and mean of 1708 at global level with associated peak with fewer extreme values on the tail at right side.

The CoVID-19 data was additional assessed to identify the manifold change points in long time series data for different phases of disease and its associated span period at global, regional and selected countries scale (Figure 4(a-b), Appendix Figure 4 to 8). Each data curve displayed 03 phases of data point change accompanied by last curve segment distance. The 1st phase characterize concealed CoVID-19 incubation interval which ultimately signify hollowness of policy linked activities to mitigate the spread. The 2nd phase signify the exponential growth time and translation into high danger for CoVID-19 spread. The 3rd change detection phase

show catastrophic influence on public health and socio-economy. The 4th and last length of curve segment represent time engaged for the disease curve flattening by implementing policies of lockdowns, face mask, social distancing and additional policy actions to restrict exponential rise in pandemic. Segment of longer length and lower height showed decrease in CoVID-19.

The data mean linked CPD shown 1st phase length of 39 - 56 days for all regions except Western Pacific with 12 days during January 22nd to March 17th, 2020) and January 22nd to February 1st, 2020, respectively. The significant rise are observed in cases for African in 3 days, 4 days for American and South East Asia regions, 5 days for East Mediterranean, in 6 days for European and 13 days for Western Pacific during 2nd phase. The 3rd phase of CPD observed in range of 2 - 4 days for all regions except for region of Eastern Mediterranean with 17 days. Last segment length observed to be 3 to 5 days except 36 days for region of Western Pacific demonstrating CoVID-19 cool down stage.

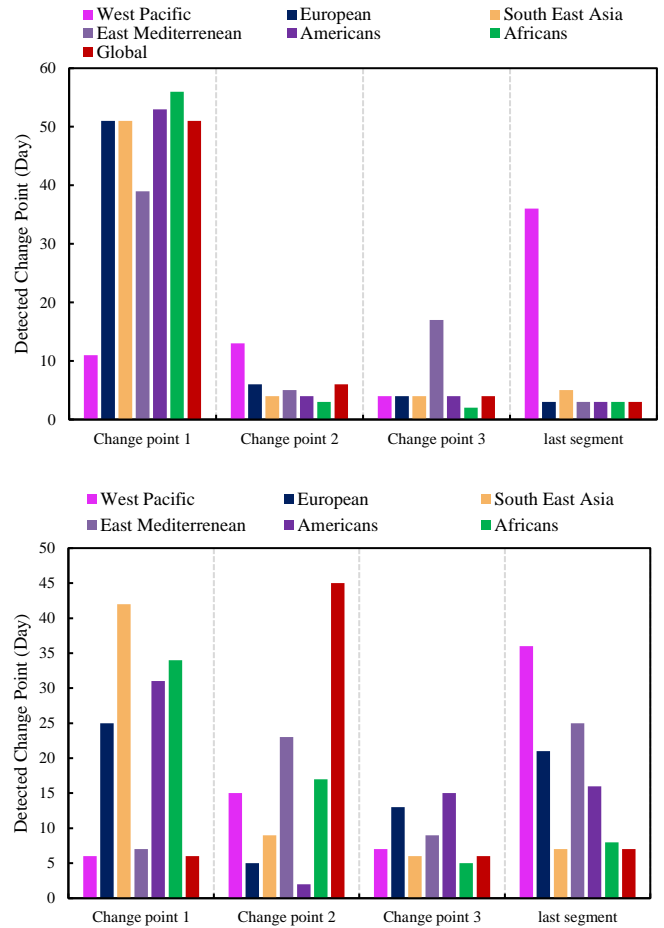
The characteristics of global pandemic data curve showed duration for first phase to last segment of the curve highly variable with 51, 6, 4 and 3 days, respectively. The data mean linked CPD disregard temporal data variance. So, CPD based on data mean and data variance provided more in-depth evidence for CoVID-19. The 1st phase length was observed shorter

for the regions of Western Pacific and East Mediterranean with 6 - 7 days during (January 22nd to 28th, 2020) followed by region of Europe with 25 days, American region with 31 days, African region with 34 days and region of South East Asia with 42 days on March 3rd, 2020). It was observed that all regions different during 2nd phase and sudden increase in cases for American region with 2 days, followed by region of Europe with 5 days, South East Asia region with 9 days, region of Western Pacific with 15 days, region of Africa with 17 days) and region of Eastern Mediterranean with 23 days. The 3rd phase of CPD showed that African region faced smallest influence due to plague with 5 days, followed by region of South East Asia region with 6 days, region of Western Pacific with 7 days, region of Eastern Mediterranean with 9 days, region of Europe with 13 days and highest for region of American with 15 days. Last segment length of CPD observed to be 7 to 25 days excluding region of Western Pacific with 36 days. The diseases curve characterised by 6, 45, 6 and 7 days from 1st phase to last segment at global level. Further details for CPD related to affected countries are available in supplementary information.

**3.3 Polynomial Regression and Projections**

The Change Point Detection based on CoVID-19 data time series although helped to comprehend the CoVID-19 pandemic by identifying different phases and sudden changes. It is also deficient in explaining temporal trends and did not provide any kind of projections for infected population. This research also attempted to assess simple but robust method of polynomial data curve fitting to compute and predict infected population (Figure 5, Appendix Figure 9 to 10).

The 2nd order polynomial was used to estimate 0.373 million infected population confirmed cases with coefficient of determination  $r^2$  of 0.87) at global level and predicted increase to 0.750 million ( $r^2 = 0.87$ ) with 2nd order, 1.900 million with coefficient of determination  $r^2$  of 0.96 with 3rd order and 4.680 million cases with  $r^2 = 0.99$  by 4th order polynomials method, by April 24, 2020. The infected population predictions were found to be in range of 0 - 0.700 million for Western Pacific region, European region with 0.500 - 2.650 million, South East Asia region with 0.006 - 0.048 million, Eastern Mediterranean region with 0.080 - 0.150 million, American region with 0.120 - 0.990 million and African region with 0.003 - 0.0230 million persons.



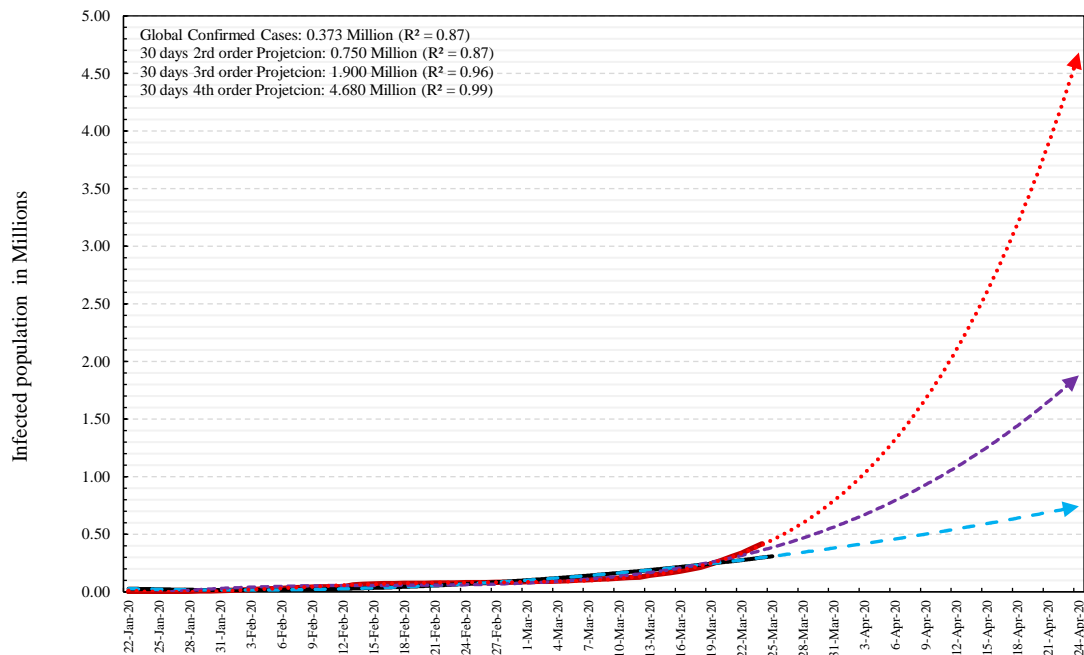
**Figure 5:** (a) The change point detection based on data mean and (b) based on data mean and variance.

<b>Table 1: Assessment of Projected Infected Population on 1st April, 2020 at Global, Regional and Selected Countries Scale. Bold are Best Models in Term of Percent Error in Projections.</b>							
Scale	Confirmed Cases (Million)	Projected Infected population (Million)			Percent Error		
	1st April 2020	2nd order	3rd order	4th order	2nd order	3rd order	4th order
Global	0.824	0.400	0.600	0.900	-51.4	-27.2	<b>9.3</b>
Western Pacific	0.106	0.095	0.090	0.140	<b>-10.7</b>	-15.4	31.6
European	0.464	0.200	0.350	0.500	-56.9	-24.6	<b>7.7</b>
South East Asia	0.007	0.003	0.006	0.008	-58.4	-16.8	<b>10.9</b>
Eastern Mediterranean	0.054	0.036	0.048	0.047	-33.7	<b>-11.6</b>	-13.4
American	0.189	0.040	0.080	0.180	-78.8	-57.6	<b>-4.6</b>
African	0.004	0.001	0.002	0.003	-75.4	-50.9	<b>-26.3</b>
China	0.083	0.070	0.065	0.120	<b>-15.3</b>	-21.3	45.2
Italy	0.106	0.080	0.120	0.140	-24.4	<b>13.4</b>	32.3
USA	0.163	0.040	0.080	0.150	-75.5	-51.0	<b>-8.1</b>
Spain	0.094	0.040	0.075	0.100	-57.6	-20.6	<b>5.9</b>
Germany	0.067	0.040	0.060	0.080	-40.6	<b>-10.9</b>	18.8
Iran	0.045	0.026	0.044	0.040	-41.7	<b>-1.4</b>	-10.3
France	0.051	0.020	0.040	0.055	-61.1	-22.3	<b>6.8</b>
Switzerland	0.016	0.009	0.016	0.024	-44.1	<b>-0.7</b>	49.0
South Korea	0.010	0.014	0.010	0.002	41.6	<b>1.1</b>	-79.8
UK	0.025	0.008	0.014	0.021	-68.2	-44.3	<b>-16.5</b>
Netherlands	0.013	0.006	0.010	0.014	-52.4	-20.6	<b>11.2</b>
Austria	0.010	0.004	0.008	0.011	-60.7	-21.4	<b>8.0</b>
Belgium	0.013	0.004	0.008	0.011	-68.7	-37.4	<b>-13.9</b>
Canada	0.008	0.002	0.004	0.007	-74.0	-48.0	<b>-9.0</b>
Pakistan	0.002	0.001	0.002	0.002	-51.0	<b>-1.9</b>	<b>-1.9</b>

Predictions for CoVID19 infected population were also carried out for selected countries which were 0.000 - 0.625 million for China, Italy with 0.180 - 0.690 million, USA with 0.115 - 0.960 million, Spain with 0.095 - 0.600 million, Germany with 0.080 to 0.480 million, Iran with 0.072 - 0.128, France with 0.050 - 0.300 million, Switzerland with 0.024 - 0.145 million, South Korea with 0.000 - 0.025 million, UK with 0.019 - 0.125 million, Netherland with 0.012 - 0.076 million, Austria with 0.012 - 0.076 million, Belgium with 0.010 - 0.061 million, Norway with 0.008 - 0.025 million, Canada with 0.006 - 0.044 million and Pakistan with 0.002 - 0.016 million persons.

The polynomial method was evaluated to conclude better order for CoVID-19 data fitting for approximation and predictions based on percentage errors (Table 1). The confirmed CoVID-19 cases of April 1st, 2020 from situation report of WHO were used to estimate the infected people predictions based on all 3 polynomial orders. The polynomial with 2nd

order was observed best for prediction of infected people for region of Western Pacific with  $r^2$  value of 0.97 and China  $r^2$  value of 0.96 with errors of -15.7 - -10.7 percent. Best predicting models for infected population of Eastern Mediterranean with  $r^2$  value of 0.99), Germany  $r^2$  value of 0.94, Iran  $r^2$  value of 0.99, Switzerland  $r^2$  value of 0.94, South Korea  $r^2$  value of 0.95, Italy  $r^2$  value of 0.88, and Pakistan  $r^2$  value of 0.90 with prediction errors of approx. -11.6 - 13.4 percent with polynomial of 3rd order fitting. However, remaining regions and countries were best predicted with polynomial model of 4th order with approx. prediction errors of -26.3 - 11.2 percent. This include region of Europe with  $r^2$  value of 0.99, South East Asia with  $r^2$  value of 0.98, American region with  $r^2$  value of 0.94 and African region with  $r^2$  value of 0.95), Spain with  $r^2$  value of 0.99, USA with  $r^2$  value of 0.95, UK with  $r^2$  value of 0.98), Netherlands with  $r^2$  value of 0.99, France with  $r^2$  value of 0.99, Belgium with  $r^2$  value of 0.99, Canada with  $r^2$  value of 0.96, Austria with  $r^2$  value of 0.99, and Pakistan with  $r^2$  value of 0.97.



**Figure 6:** Global projected population based on different polynomial orders.

The CoVID-19 infected population predictions based polynomial model fitting with different orders improved our understanding about pandemic status and possible association with socio-economic decisions and policy making. The polynomial predictions with 2nd order associated with region of Western Pacific especially China clarified that CoVID-19 wave number one passed its maxima and diminished to lowest. While, all other regions and selected countries were best predicted with polynomial models of 3rd order and reached closer to maxima; mainly due to late actions on devised policies or horizontal spread over indigenous transmission. The polynomial models with 4th order denotes where disease transmission was at initial phase and projected to propagate exponentially; demanding actions of strict policies to restrict epidemic.

#### 4. CONCLUSIONS

Key outcomes remained that CoVID-19 global pandemic exhibited analogous level of infection and recovery rate up to 9.0 percent at daily scale and to smooth the plague curve. The statistical evaluations based on PDF's, skewness values and kurtosis measures helped to explain disease curve shapes and peak height. The analysis based change point detection was conducted to determine different distinctive phases of CoVID-19 pandemic through observed infected population. The China was the first country, where 1st pandemic wave had ended. Remaining countries displayed that pandemic was still active and time of maxima, size of infected residents will be determined by national action policies effectiveness well supported community role like wearing of mask and social distancing, enhanced capacity of health system, community food access and availability of medicines and others.

Polynomial models were used because of its ability in explaining CoVID-19 nonlinearity. The polynomial model of 2nd order was observed better for China as CoVID-19 was already crossed its peak. However, disease at initial to peak stages polynomials models of 3rd and 4th orders were best

to estimate exponential increasing infected cases and horizontal spread of disease. Infected population predictions based on polynomial models were subject to high error risk due to highly variable nature of CoVID-19 daily situation and demands for 3-7 days repeated projections to better control the disease. Limitation to current study could be that it does not take into consideration the mathematical modeling that can be helpful in taking into account the transmission dynamics of the CoVID-19. Further, vaccination impact on the projections was not included as it was not developed at that time. In future, study will be expanded by extending the data time series and integration of geographic information systems to embed the spatial context.

#### ACKNOWLEDGEMENT

Author acknowledges the availability of good quality daily CoVID-19 data from World Health Organization (WHO) for this research.

#### REFERENCES

- Abdulrahman, I., 2023. Sim COVID: Open-source simulation programs for the COVID-19 outbreak. *SN Computer Science*, 4 (1), Pp. 20.
- Akhtar, I.H., 2020. Understanding the CoVID-19 pandemic curve through statistical approach. *medRxiv*, Pp. 2020-04.
- Akhtar, I.H., and Athar, H., 2019. Contribution of changing precipitation and climatic oscillations in explaining variability of water extents of large reservoirs in Pakistan. *Scientific Reports*, 9 (1), Pp. 1-14.
- Akhtar, I.H., and Athar, H., 2020. Water supply and effective rainfall impacts on major crops across irrigated areas of Punjab, Pakistan. *Theoretical and Applied Climatology*, 142, Pp. 1097-1116.

- Akhtar, I.H., and Athar, H., 2021. Climate-related inter-annual variability and long-term influence on wheat yield across canal-irrigated areas of Punjab, Pakistan. *Theoretical and Applied Climatology*, 143, Pp. 1195–1211.
- Anderson, R.M., 1991. Discussion: The Kermack-McKendrick epidemic threshold theorem. *Bulletin of Mathematical Biology*, 53 (1-2), Pp. 1–32.
- Bedford, J., Farrar, J., Ihekweazu, C., Kang, G., Koopmans, M., and J., 2019. Nkengasong, A new twenty-first century science for effective epidemic response. *Nature*, 575, Pp. 130–136.
- Bhattacharjee, S., 2020. Statistical investigation of relationship between spread of coronavirus disease (COVID-19) and environmental factors based on study of four mostly affected places of China and five mostly affected places of Italy. *ArXiv preprint, arXiv:2003.11277*.
- Casella, F., 2020. Can the COVID-19 epidemic be controlled on the basis of daily test reports? *IEEE Control Systems Letters*, 5, Pp. 1079–1084.
- Chissom, B.S., 1970. Interpretation of the kurtosis statistic. *The American Statistician*, 24, Pp. 19–22.
- Cox, D., Kartsonaki, C., Keogh, R., 2018. Big data: Some statistical issues. *Statistics and Probability Letters*, 136, Pp. 111–115.
- Cui, J., Li, F., and Shi, Z.L., 2019. Origin and evolution of pathogenic corona viruses. *Nature reviews Microbiology*, 17, Pp. 181–192.
- Da'ar, O.B., and Ahmed, A.E., 2018. Underlying trend, seasonality, prediction, forecasting and the contribution of risk factors: an analysis of globally reported cases of Middle East Respiratory Syndrome Coronavirus. *Epidemiology and Infection*, 146, Pp. 1343–1349.
- Deb, S., and Majumdar, M., 2020. A time series method to analyze incidence pattern and estimate reproduction number of COVID-19. *ArXiv preprint, arXiv:2003.10655*.
- Doane, D.P., and Lori, E.S., 2011. Measuring skewness: a forgotten statistic? *Journal of statistics education*, Pp. 19, <https://doi:10.1080/10691898.2011.11889611>.
- Dong, E., Du, H., and Gardner, L., 2020. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, [https://doi:10.1016/S1473-3099\(20\)30120-1](https://doi:10.1016/S1473-3099(20)30120-1).
- Duong, T., 2007. ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R. *Journal of Statistical Software*, 21, Pp. 1–16.
- Edwards, A.W.F., and Cavalli-Sforza, L.L., 1965. A method for cluster analysis. *Biometrics*, 21, Pp. 362–375.
- Eissa, M.E., Rashed, E.R., and Eissa, D.E., 2022. Modeling of COVID-19 Major Outbreak Wave Through Statistical Software: Quantitative Risk Evaluation and Description Analysis. *ESTÜDAM Halk Sağlığı Dergisi*, 7 (1), Pp. 145–61.
- Gaeta, G., 2020. A simple SIR model with a large set of asymptomatic infective. *ArXiv preprint, arXiv:2003.08720*.
- Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A., and Colaneri, M., 2020. A SIDARTHE model of COVID-19 epidemic in Italy. *ArXiv preprint, arXiv:2003.09861*.
- Giuliani, D., Dickson, M.M., Espa, G., and Santi, F., 2020. Modelling and predicting the spatio-temporal spread of COVID-19 in Italy. *BMC Infectious Diseases*, 20 (1), Pp. 10.
- Haider, N., Yavlinsky, A., Simons, D., Osman, A.Y., Ntoumi, F., Zumla, A., and Kock, R., 2020. Passengers' destinations from China: low risk of novel coronavirus (2019-nCoV) transmission into Africa and South America. *Epidemiology & Infection*, 148, <https://doi:10.1017/S0950268820000424>.
- Hellewell, J., Abbott, S., Gimma, A., Bosse, N.I., Jarvis, C.I., Russell, T.W., Munday, J.D., Kucharski, A.J., Edmunds, W.J., Sun, F., and Flasche, S., 2020. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health*, 8, Pp. 488–496.
- Hembram, K.P.S.S., and Kumar, J., 2021. Epidemiological study of novel coronavirus (COVID-19): macroscopic and microscopic analysis. *Int J Commun Med Public Health*, 8, Pp. 1364.
- Hill, A.B., 1965. The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine*, 58, Pp. 295–300.
- Hinkley, D.V., 1970. Inference about the change point in a sequence of random variables. *Biometrika*, 57, Pp. 1–17.
- Inglesby, T.V., Nuzzo, J.B., O'Toole, T., and Henderson, D.A., 2006. Disease mitigation measures in the control of pandemic influenza. *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science*, 4, Pp. 366–375.
- Jayaraj, V.J., Chong, D.W.Q., Wan, K.S., Hairi, N.N., Bhoo-Pathy, N., Rampal, S., and Ng, C.W., 2023. Estimating excess mortalities due to the COVID-19 pandemic in Malaysia between January 2020 and September 2021. *Scientific Reports*, 13 (1), Pp. 86.
- Jones, D.S., 2020. History in a crisis lesson for Covid-19. *New England Journal of Medicine*, 382, Pp. 1681–1683. <https://doi:10.1056/NEJMp2004361>.
- Killick, R., and Eckley, I., 2014. Change point: An R package for change point analysis. *Journal of statistical software*, 58, Pp. 1–19.
- Koo, J.R., Cook, A.R., Park, M., Sun, Y., Sun, H., Lim, J.T., Tam, C., Dickens, B.L., 2020. Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. *Lancet Infectious Diseases*, 20 (6), Pp. 678–688.
- Lee, H., Jang, G., and Cho, G., 2022. Forecasting COVID-19 cases by assessing control-intervention effects in Republic of Korea: a statistical modeling approach. *Alexandria Engineering Journal*, 61 (11), Pp. 9203–9217.
- Morens, D.M., and Fauci, A.S., 2007. The 1918 influenza pandemic: insights for the 21st century. *The Journal of infectious diseases*, 195, Pp. 1018–1028.
- Nadim, S.S., Ghosh, I., and Chattopadhyay, J., 2021. Short-term predictions and prevention strategies for COVID-19: a model-based study. *Applied mathematics and computation*, 404, Pp. 126251.
- Ospina, R., Leite, A., Ferraz, C., Magalhães, A., and Leiva, V., 2022. Data-driven tools for assessing and combating COVID-19 outbreaks in Brazil based on analytics and statistical methods. *Signa Vitae*, 18 (3), Pp. 18–32.
- Page, E.S., 1954. Continuous inspection schemes. *Biometrika*, 41, Pp. 100–115.
- Perperoglou, A., Sauerbrei, W., Abrahamowicz, M., and Schmid, M., 2019. A review of spline function procedures in R. *BMC medical research methodology*, 19, Pp. 1–16.
- Picard, F., Robin, S., Lavielle, M., Vaisse, C., and Daudin, J., J., 2005. A statistical approach for array CGH data analysis. *BMC bioinformatics*, 6, Pp. 1–14.
- Rao, A.S.S., and Vazquez, J.A., 2020. Identification of COVID-19 can be quicker through Artificial Intelligence framework using a mobile phone based survey in the populations when cities/towns are under quarantine. *Infection Control and Hospital Epidemiology*, 4, Pp. 826–830, <https://doi:10.1017/ice.2020.61>.
- Sattenspiel, L., 1987. Population structure and the spread of disease. *Human Biology*, 59, Pp. 411–438.
- Sen, A., and Srivastava, M.S., 1975. On tests for detecting change in meaning. *The annals of statistics*, 3, Pp. 98–108.
- Shi, H., Han, X., Jiang, N., Cao, Y., Alwalid, O., Gu, J., Fan, Y., and Zheng, C., 2020. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. *The Lancet infectious diseases*, 20, Pp. 425–434.
- Smith, D., and Moore, L., 2004. The SIR model for spread of disease: the differential equation model. *Loci. (Originally convergence.)*, <https://www.maa.org/press/periodicals/loci/joma/>.
- Taimoor, M., Ali, S., Shah, I., and Muwanika, F.R., 2022. COVID-19 pandemic



- data modeling in Pakistan using time-series SIR. *Computational and Mathematical Methods in Medicine*, 14, Pp. 6001876. <https://doi.org/10.1155/2022/2001876>
- The Lancet, 2020. Emerging understandings of 2019-nCoV. *Lancet*, 395, Pp. 311, [https://doi.org/10.1016/S0140-6736\(20\)30186-0](https://doi.org/10.1016/S0140-6736(20)30186-0).
- Tosto, S., 2022. Statistical investigation on covid 19 epidemics. *Aeronautics and Aerospace Open Access Journal*, 6 (3), Pp. 83–88.
- Wan, Y., Shang, J., Graham, R., Baric, R.S., and Li, F., 2020. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *Journal of Virology*, 94 (7), Pp. e00127-20.
- Wand, M.P., and Jones, M.C., 1995. *Kernel Smoothing*. Chapman and Hall, London.
- Wang, P., Liu, H., Zheng, X., and Ma, R., 2023. A new method for spatio-temporal transmission prediction of COVID-19. *Chaos, Solitons and Fractals*, 167, Pp. 112996.
- WHO., 2020. Novel Coronavirus (2019-nCoV) situation reports, <https://www.who.int/>.
- Wilks, D.S., 2011. *Statistical methods in the atmospheric sciences*, 3rd Edition. Academic Press, USA.
- Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z., W., Tian, J.H., Pei, Y.Y., and Yuan, M.L., 2020. A new coronavirus associated with human respiratory disease in China. *Nature*, 579, Pp. 265–269.
- Zhao, S., Cao, P., Chong, M.K., Gao, D., Lou, Y., Ran, J., Wang, K., Wang, W., Yang, L., He, D., and Wang, M.H., 2020. The time-varying serial interval of the coronavirus disease (COVID-19) and its gender-specific difference: a data-driven analysis using public surveillance data in Hong Kong and Shenzhen, China from January 10 to February 15, 2020. *Infection control & hospital epidemiology*, 10, Pp. 1–8, <https://doi.org/10.1017/ice.2020.64>.
- Zhong, L., Mu, L., Li, J., Wang, J., Yin, Z., Liu, D., 2020. Early Prediction of the 2019 Novel Coronavirus Outbreak in the Mainland China Based on Simple Mathematical Model. *IEEE Access*, 8, Pp. 51761-51769. <https://doi.org/10.1109/ACCESS>.
- Zrieq, R., Kamel, S., Boubaker, S., Algahtani, F.D., Alzain, M.A., Alshammari, F., Alshammari, F.S., Aldhmadi, B.K., Atique, S., Al-Najjar, M.A. and Villareal, S.C., 2022, September. Time-Series Analysis and Healthcare Implications of COVID-19 Pandemic in Saudi Arabia. *Healthcare*, 10 (10), Pp. 1874.

