JMSR
Journal of Medical
and Scientific Research

ORIGINAL RESEARCH

Spatial modeling for COVID-19 analysis: An Indian case study

Muralikrishna Iyyanki[1], Jayanthi Prisilla[2,*], and Sudarshan Kandle[3]

[1]Former Dr. Raja Ramanna Distinguished Fellow DRDO and Director R&D JNT University, Hyderabad, Telangana, India
[2]The Airport Authority of India, Shamshabad, Hyderabad, Telangana-500409, India
[3]Department of Geography, Osmania University,Tarnaka, Hyderabad, Telangana-500007, India

Abstract

The coronavirus disease 2019 (COVID-19) outbreak in India from January 31, 2020, onwards to June 15, 2020, has reached confirmed cases over 3,32,424 that are being reported. The aim of this study is to predict and explore the spatial distribution of COVID-19 data of India using three models – geographical weighted regression (GWR), generalized linear regression (GLR), and ordinary least square (OLS). In this paper, the swift rise in COVID-19 cases is experiential after the lockdown period. This is explored using ArcGIS on the confirmed case of June 15, 2020, as the response with the explanatory of COVID-19 cases, i.e March 15, 2020, April 7, April 12, May 12, and June 1, 2020. The confirmed cases of the dataset is classified into three cases ie. case-1: June 15, 2020, vs March 15 and April 7, 2020; case-2: June 15, 2020 vs April 12, May 12 and June 1, 2020; and case-3: June 15, 2020 Vs all dates mentioned in discussion Hence, the prediction using GWR gave the much closer values for June 16, 2020. AICc of GWR (618.9038) was found to have the minimum value over GLR and OLS models. The day-wise increase and samples tested per day in twelve different states is analyzed using STATA. The number of testing varies with states to states, depending on the population and testing labs available. The percentage for each slope is achieved as m1 (-5.714 %), m2 (39.393%), m3 (6.521%) and m4 (46.938%).

***Keywords**: COVID-19; GIS; spatial data; spatial models; testing samples*

***Corresponding author:** Prisilla Jayanthi, The Airport Authority of India, Shamshabad, Hyderabad, Telangana 500409, India. Email: prisillaj28@gmail.com

Received 11 August 2019; *Revised* 26 October 2020; *Accepted* 19 November 2020; *Published* 27 November 2020

Citation: Iyyanki M, Prisilla J, Kandle S. Spatial modeling for COVID-19 analysis: An Indian case study. J Med Sci Res. 2020; 8(S1):19-32. DOI: http://dx.doi.org/10.17727/JMSR.2020/8S1-3

1. Introduction

In India, the coronavirus disease 2019 (COVID-19) is the global pandemic of coronavirus share caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The first observed case of COVID-19 in India was initiated from China on January 30, 2020. This virus has spread rapidly across the whole country, especially Maharashtra with the highest confirmed cases of 107958 (June 15, 2020). COVID -19 has a significant correlation with air quality, average, and minimum temperature [1]. The two transmission mode of corona is respiratory and contact. The sanitation and hygienic environments are crucial to protect human health during this infectious COVID-19 outbreak. Ensuring decent and frequent hand wash practices in communities, homes, and health care will help prevent and reduce man-to-man transmission of the COVID-19 virus was avowed by World Health Organization (WHO) [2]. The physical examination of the patients was found to have dry mucous membranes, difficulty breathing, sore throat, headache, or cough [3]. All these lead to a lockdown of many countries, including India from March 25, 2020, to May 31, 2020, in four phases. In few areas of containment zone, the lockdown is extended up to June 30, 2020, as fifth phase. In this study, the spatial data i.e confirmed cases and testing samples, are focused. The confirmed cases are focused on understanding the rate of increase per day in every state and thereby in India. Also, the testing sample per day in every state is examined.

Few studies on geographical weighted regression (GWR) were explored for this study, and implemented on Indian COVID-19 data. Mollalo for COVID-19 data in the US has performed different models on the dataset, which included thirty-five environmental variables, socioeconomic, behavioral, topographic and demographic factors. The five different models used were three global models, namely ordinary least square (OLS), SLM and SEM, and two local models, namely GWR and multiscale GWR (MGWR). The results of MGWR achieved the highest goodness-of-fit with the most parsimonious model compared to others. The spatial variability of MGWR in different countries can reflect different behavior of COVID-19 cases in response to the explanatory variables [4].

Wang performed GWR to examine the relationship between the index of frequency of extreme precipitation and other climatic extreme indices in China that includes the frequency of warm days, warm nights, cold days, and cold nights. Based on statistical tests, the regression relationship was observed to be significant between spatial non-stationarity and explanatory variables that exhibited significant spatial inconsistency. GWR was implemented in a case of ecological inference to solve the problems related to the inference of the individual [5]. Calvo & Escolar proposed GWR approach for solving complications of spatial aggregation bias and spatial autocorrelation that affect all well-known approaches of ecological inference. This estimation process can theoretically and intuitively compute, showing that GWR approach to Goodman and King's Ecological Inference methods results in unbiased and consistent local estimates of ecological data that reveal extreme spatial heterogeneity [6]. GWR on data of house price varying with both power and rotation parameters to generate different Minkowski distances, the study proved that the local collinearity can be both negatively and positively affected by distance metric choice. The results indicate that distance metric choice can provide a useful extra tuning component to address local collinearity issues in spatially varying coefficient modelling and helps to understand the interaction of distance metric and collinearity can provide insight into the nature and structure of the data relationships [7].

Franch-Pardo carried out an assessment of sixty three scientific articles on geospatial and spatial-statistical analysis of COVID-19. The study is grouped into the categories of disease mapping: spatiotemporal analysis, health and social geography, environmental variables, data mining, and web-based mapping. It was clarified that the spatiotemporal dynamics of COVID-19 needs very strong decision making, planning and community action. Also, it emphasized that the challenges from an interdisciplinary perspective with proactive planning, international solidarity and a global perspective needs to be addressed to fight COVID-19 [8]. Gupta used long-term climatic data of air temperature (V1), rainfall (V1), actual evapotranspiration (V1), solar radiation (V1), specific humidity, wind speed with topographic altitude and density of population at the regional point to examine the spatial association with the quantity of COVID-19 infections. Their results proved Variable Importance of Projection through PLS technique that had very higher significance over all V1's [9].

Boulos & Geraghty (2020) discusses about the disease mapping and the social media reactions for disease spread, predictive risk mapping using population travel data, tracing and mapping super-spreader trajectories and contacts across space and time. The study is how GIS and mapping dashboards can support the fight against infectious disease outbreaks and epidemics [10]. Krishnakumar & Rana gives good insights to make effective approach to culminate the world threat COVID-19 in India [11]. Pulla (2020) expresses that the transmission of COVID-19 by asymptomatic people would reduce the effectiveness of airport screening and quarantine measures. It was communicated that India would have confirmed cases of COVID-19 between around 100 000 and 1.3 million by the middle of May if the virus continues to spread at its current rate [12].

2. Data and methodology

The study includes the spatial models, namely ordinary least square (OLS), geographical weighted regression (GWR), and generalized linear regression (GLR). The details of each models is discussed in Spatial Regression Models section. The day to day data was collected from Ministry of health and family welfare (Table 1) and analyzed using ArcGIS Pro with spatial models. The samples for testing data for state-wise and all over India were obtained from Statista and ICMR websites. The testing data were found to vary based on testing labs available in each state. Total number of samples tested as on June 15, 2020 in few states such as TN (729002), MH (671348), RJ (609296) and AP (567375) were updated when compared with the confirmed cases in TN (44661), MH (107958), RJ (12694) and AP (6163). The time series graph for day wise increase in India and twelve different states was obtained using STATA 12 IC. The two way graph for sample tested and confirmed case for states was performed on STATA 12 IC. The study includes the increase of COVID-19 confirmed cases after the lockdown period and is analyzed using ArcGIS on the confirmed case of June 15, 2020 as the response variable with the explanatory variables of COVID-19 cases i.e. March 15, 2020, April 7, April 12, May 12 and June 1, 2020. Here, for this study, the confirmed cases of dataset is classified into three cases ie. case-1: June 15, 2020 Vs March 15 and April 7, 2020; case-2: June 15, 2020 Vs April 12, May 12 and June 1, 2020; and case-3: June 15, 2020 Vs March 15, April 7, April 12, May 12 and June 1, 2020.

2.1 Spatial regression models

2.1.1 Ordinary least square

Ward & Gleditsch discusses about OLS as a linear regression approach that examines the relationships between dependent variable and a set of explanatory variables and is represented with the following notation (eq. 1)

$$y_i = \beta_0 + x_i \beta + \epsilon_i \qquad (1)$$

where i represents any country, y_i is the confirmed cases (dependent variable), the intercept of the model (β_0), the vector of selected explanatory variables (x_i), the vector of regression coefficients (β), and random error term (ε_i) [13]. Based on the nature of the spatial dependence, OLS will be either incompetent with incorrect standard errors or biased and inconsistent [14]. If spatial dependence among the data exists, then it violates assumptions about the error term [15, 16].

2.1.2 Generalized linear regression

GLR is a regression model used to generate predictions or to model a dependent variable in relation to a set of explanatory variables. Its prediction can be used to examine and quantifies relationships among features. The tool is used to fit continuous (OLS), binary (logistic), and count (Poisson) models. A count model assumes that the mean and variance of the dependent variable are equal, and moreover, the values of the dependent variable cannot be negative or contain decimals. The notation of GLR is as follows in eq. 2.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + e_1 \quad (2)$$

where β_0 is the intercept, β_1 and β_2 are the slope and coefficient of the explanatory variables in regressions with x_1 x_2 and, ...x_n, respectively. The term e_i is the error terms, and *y* is the dependent variable [17].

Table 1: Covid- 19 data as on June 15, 2020.

S. No.	*State*	*Confirmed cases*	*Recovered*	*Death*
1	AN Islands	38	33	0
2	AP	6163	3314	84
3	AR	91	7	0
4	AS	4049	1960	8
5	BR	6470	4170	39
6	CH	352	293	5
7	CG	1662	763	8
8	DD	36	2	0
9	DL	41182	15823	1327
10	GA	564	74	0
11	GJ	23544	16325	1477
12	HR	7208	3003	88
13	HP	518	337	7

14	JK	5041	2389	59
15	JH	1745	905	8
16	KA	7000	3955	86
17	KL	2461	1102	19
18	LA	549	80	1
19	MP	10802	7677	459
20	MH	107958	50978	3950
21	MN	458	91	0
22	ML	44	25	1
23	MZ	112	1	0
24	NL	168	88	0
25	OD	3909	2708	11
26	PY	194	91	5
27	PB	3140	2356	67
28	RJ	12694	9566	292
29	SK	68	4	0
30	TN	44661	24547	435
31	TS	4974	2377	185
32	TR	1076	315	1
33	UK	1819	1111	24
34	UP	13615	8268	399
35	WB	11087	5060	475
Total		332424	169798	9520

2.1.3 Geographically weighted regression

GWR is a spatial techniques mostly used in geography and many other disciplines. GWR is a local model of the variable to predict by fitting a regression equation to each feature in the dataset. It should be noted that GWR is not an appropriate method for small datasets and does not work with multipoint data. The notation of GWR is given in eq. 3 [4].

$$y_i = \beta_{i0} + \sum_{j=1}^{m} \beta_{ij} X_{ij} + \varepsilon_i, \; i = 1, 2, \ldots, n \qquad (3)$$

where i is a country, y_i is the value for the confirm cases, the intercept (β_{i0}), the jth regression parameter (β_{ij}), X_{ij} is the value of the jth explanatory parameter, and ε_i is a random error term.

3. Findings and results

The models are generated on the datasets obtained from https://www.mohfw.gov.in/ and the analysis was performed on ArcGIS Pro. The results of OLS model (Figure 1) summaries the coefficient, T-statistic and P-value along with VIFs on explanatory variables assumed (Tables 2, 3 and 4); the selected variables have relatively low multi-collinearity since the Variance Inflation Factor (VIFs) for all of explanatory variables were positively associated with confirmed cases ($p< 0.01$). The p-value for Con-June-1 (0.0000) is much better with VIF (66.16) over the other explanatory in case-1, similarly in case-II, p-value of Con-Apr-7(0.00001) is good fit; and Con-June-1 (0.0000) in case-III.

Table 2: Summary of OLS model on explanatory variables –Case-1.

Var	*Coeff.*	*T-statistic*	*P-Value*	*VIF*
Intercept	189.2513	0.6132	0.5443	---
Con-Mar-15	-173.9137	-2.7939	0.0089	3.1445
Con-Apr-7	23.9996	3.1082	0.0041	15.8673
Con-Apr-12	-0.9229	-0.2325	0.8177	30.5448
Con-May-12	-2.0101	-3.9724	0.0004	72.7011
Con-June-1	2.2400	12.9313	0.0000	66.1686

Table 3: Summary of OLS model on explanatory variables – Case –II.

Var	*Coeff.*	*T-statistic*	*P-Value*	*VIF*
Intercept	-1320.4522	-0.5970	0.5546	----
Con-Mar-15	587.1253	1.6602	0.1063	1.8876
Con-Apr-7	101.1918	5.1819	0.00001	1.8876

Table 4: Summary of OLS model on explanatory variables – Case –III.

Var	Coeff.	T-statistic	P-Value	VIF
Intercept	187.4760	0.5468	0.5883	----
Con-Apr-12	10.04214	5.0953	0.00002	5.8852
Con-May-12	-2.2176	-4.0508	0.0003	6.4994
Con-June-1	2.1562	11.6117	0.0000	59.4165

GLR model (Figure 2) with the summary of the three different models is shown in the Tables 5, 6 and 7. The z-score for the intercept is 1720.1645, Con-Mar-15 (-28.6294), Con-Apr-7 (58.2404), Con-Apr-12 (52.3568), Con-Mar-12 (166.8258) and Con-June-1 (-143.532) in table 5. Z-scores are standard deviations. Both z-scores and p-values are associated with the standard normal distribution. In table 6 the z-score for intercept (2199.1819), Con-Mar-15 (51.5807), and Con-Apr-7 (575.7223) and in table 7, the z-score for intercept (1769.2154), Con-Apr-12 (359.7744), Con-May-12 (168.4036), and Con-June-1 (-180.429).

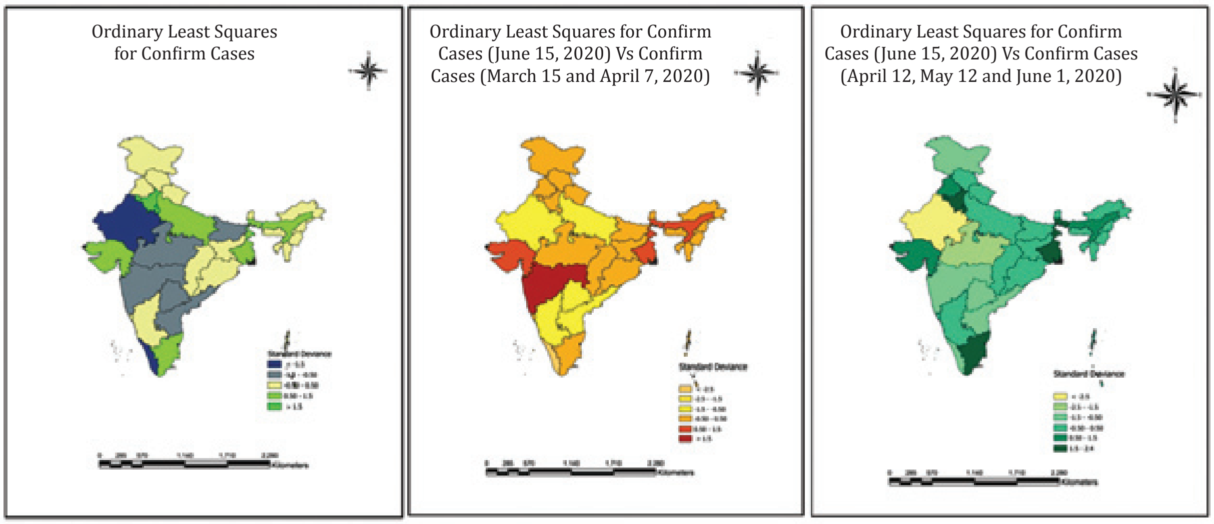

Figure 1: OLS model for confirmed cases in India.

Table 5: Summary of GLR model on explanatory variables – Case-I.

Var.	*Coeff.*	*SE*	*z-score*	*P-value*	*VIF*
Intercept	7.5277	0.0043	1720.1645	0.0000	--
Con-Mar-15	-0.0117	0.0004	-28.6294	0.0000	3.1445
Con-Apr-7	0.0030	0.00005	58.2404	0.0000	15.8673
Con-Apr-12	0.0014	0.00003	52.3568	0.0000	30.5448
Con-May-12	0.0004	0.000003	166.8258	0.0000	72.7011
Con-June-1	-0.0001	0.000001	-143.532	0.0000	66.1686

Table 6: Summary of GLR model on explanatory variables –Case-II.

Var.	*Coeff.*	*SE*	*z-score*	*P-value*	*VIF*
Intercept	7.7134	0.0035	2199.1819	0.0000	--
Con-Mar-15	0.0082	0.0001	51.5807	0.0000	1.8876
Con-Apr-7	0.0070	0.00001	575.7223	0.0000	1.8876

Table 7: Summary of GLR model on explanatory variables –Case-III.

Var.	Coeff.	SE	z-score	P-value	VIF
Intercept	7.5141	0.0042	1769.2154	0.0000	--
Con-Apr-12	0.0030	0.000009	359.7744	0.0000	5.8852
Con-May-12	0.0004	0.000003	168.4036	0.0000	66.4994
Con-June-1	-0.0002	0.000001	-180.429	0.0000	59.4165

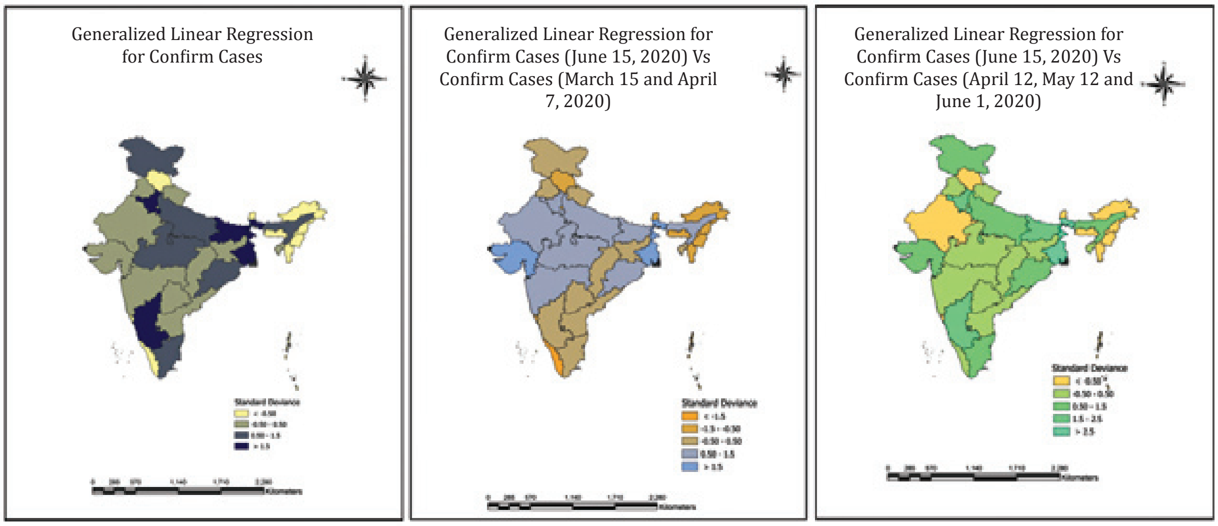

Figure 2: GLR model for confirmed cases in India.

GWR model (Figure 3) with the summary of the three different models (Tables 8, 9 and 10); the model is described based on the various factors such as Sigma square and Sigma Square MLE and DF. Finally, sigma-squared is used for AICc computations. From the tables obtained R^2 (0.9983) in Case-I is much closer to 1 over other cases i.e, R^2 (0.7804) in Case-II and R^2 (0.9939) in Case-III. And based on AICc and R^2 values, the model is coined as better model [18].

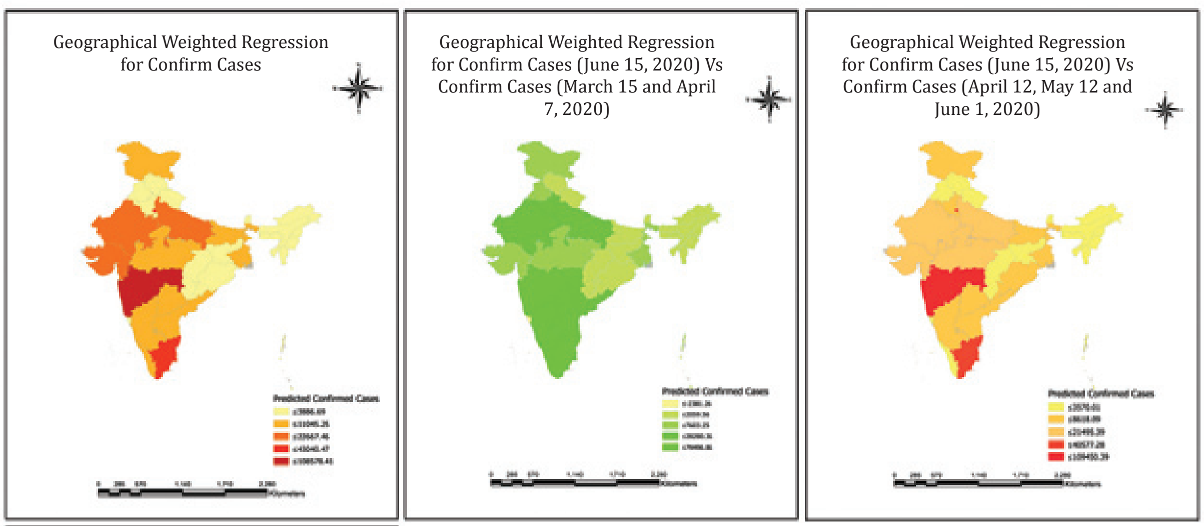

Figure 3: GWR model for confirmed cases in India.

Table 8: Summary of GWR model on explanatory variables –Case-I.

Dependent	*Confirm*
Explanatory	Con-Mar-15 Con-Apr-7 Con-Apr-12 Con-May-12 Con-June-1
R^2	0.9983
AICc	618.9038
Sigma-Sq.	977057.1123
Sigma-Sq. MLE	644959.6972
DF	23.7638

Table 9: Summary of GWR model on explanatory variables –Case-II.

Dependent	*Confirm*
Explanatory	Con-Mar-15 Con-Apr-7
R^2	0.7804
AICc	778.8805
Sigma-Sq.	108456715.0911
Sigma-Sq. MLE	84431955.5069
DF	28.0255

Table 10: Summary of GWR model on explanatory variables –Case-III.

Dependent	*Confirm*
Explanatory	Con-Apr-12 Con-May-12 Con-June-1
R^2	0.9939
AICc	651.0711
Sigma-Sq.	3047501.0060
Sigma-Sq. MLE	2341930.3485
DF	27.6651

The relationship charts (Figure 4, 5 and 6) between variables ie dependent and explanatory. The R-squared obtained (Figure 4) GLR (R^2 = 0.99) and GWR (R^2 = 0.9983) similarly (Figure 5) GLR (R^2 = 0.68) and GWR (R^2 = 0.7804) and (Figure 6) GLR (R^2 = 0.99) and GWR (R^2 = 0.9939) respectively. Deviance residual chart for GLR and GWR is shown in Figure 7, 8 and 9.

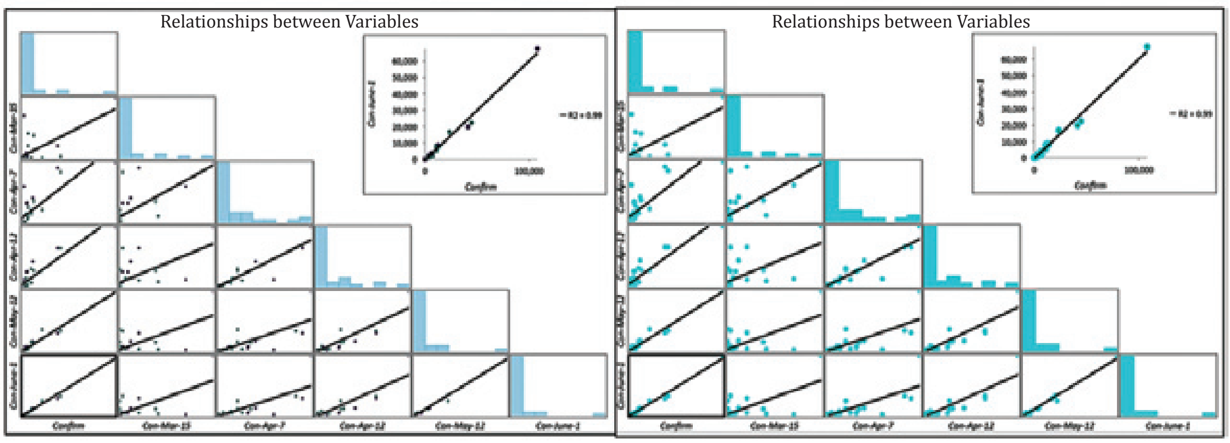

Figure 4: Case- I- Relationships graph of GLR (Left) and GWR (Right).

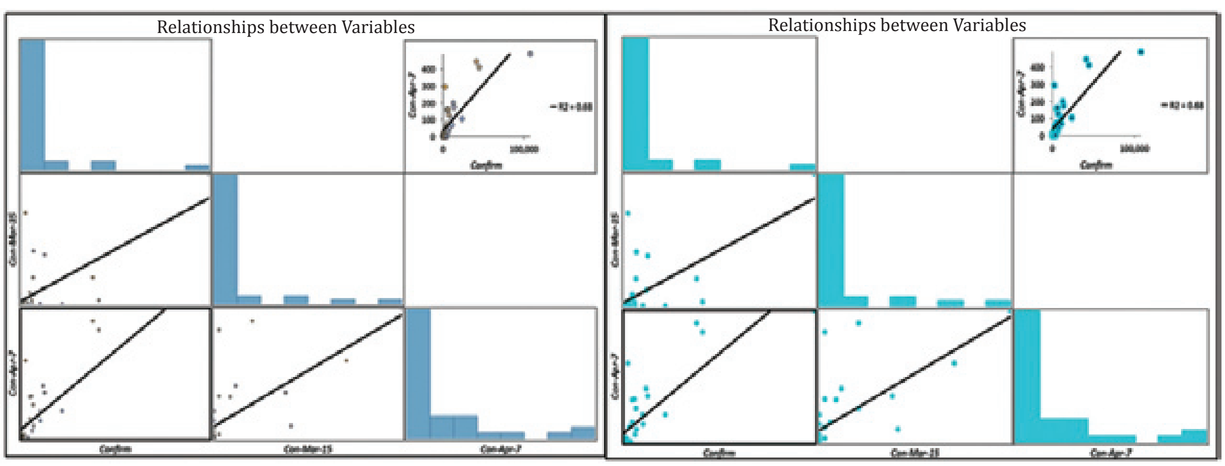

Figure 5: Case- II- Relationships graph of GLR (Left) and GWR (Right).

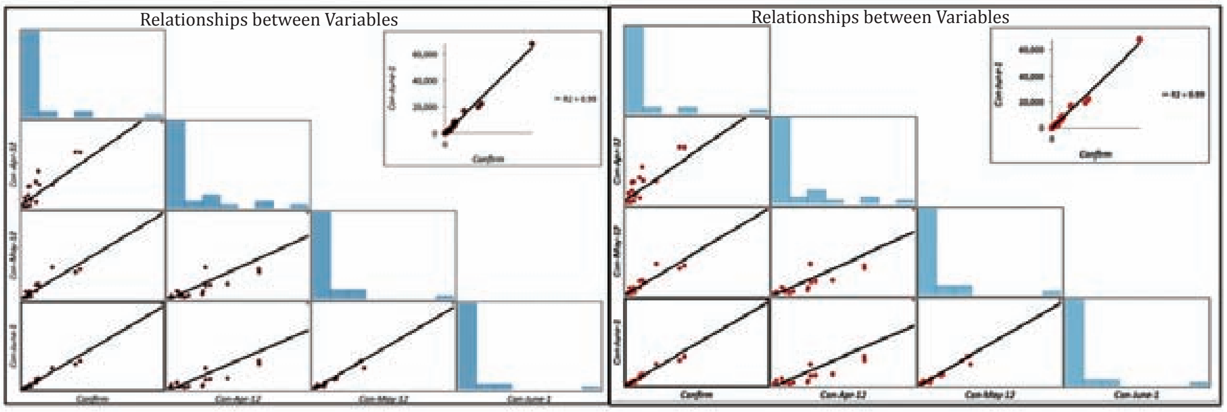

Figure 6: Case- III -Relationships graph of GLR (Left) and GWR (Right).

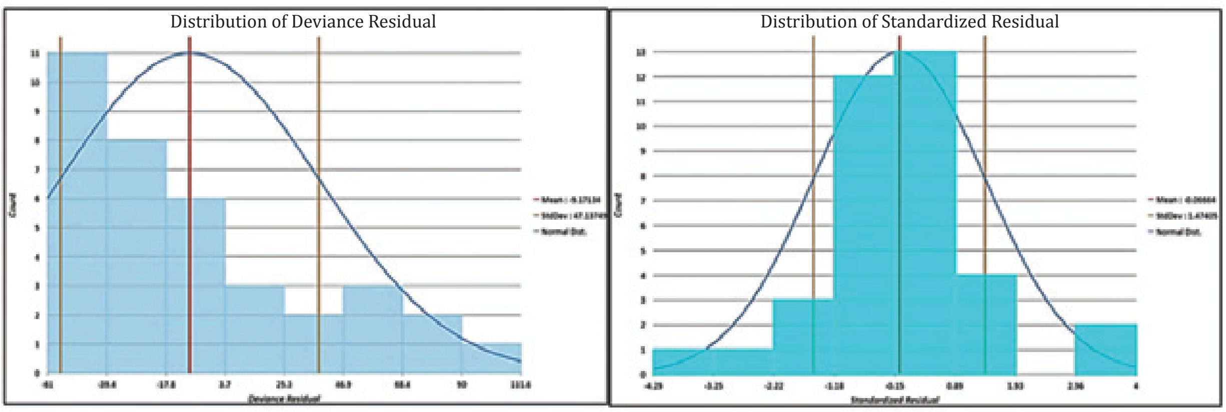

Figure 7: Case- I- Distribution of Deviance Residual graph for GLR (Left) and GWR (Right).

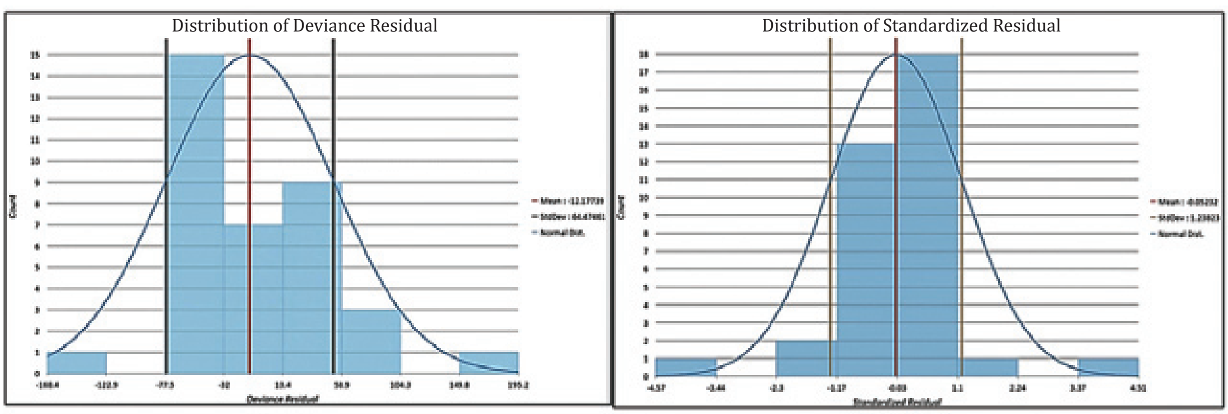

Figure 8: Case- II- Distribution of Deviance Residual graph for GLR (left) and GWR (Right).

On comparing the three models (Table 11) the values of AICc and Adj R^2, implies that the measure of model performance of regression models can be understood with AICs value. The model with the lower AICc value provides a better fit to the observed data. On examining table 11, GWR model (618.9038) is better over GLR (81132) and OLS (641.192). Adj R-sq value varies from 0.0 to 1.0 and Adj-R-sq of GWR is much nearer to 1 compared to OLS and GLR model. Hence, GWR model is the better fit model for the COVID-19 data in this study.

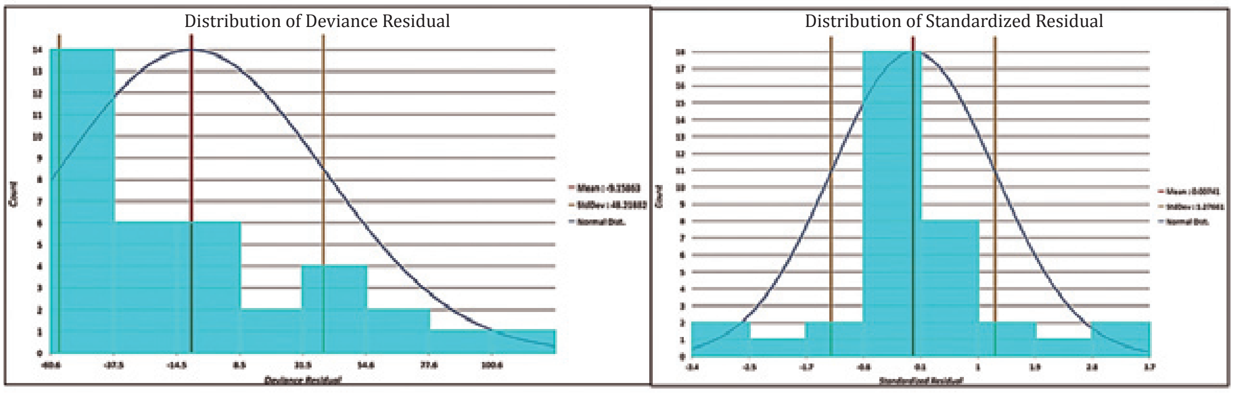

Figure 9: Case- III- Distribution of Deviance Residual graph for GLR (left) and GWR (Right).

Table 11: Measure of model fit/ performance for OLS, GWR and GLR in modelling of COVID-19 confirm cases in India.

Criterion	*OLS*			*GLR*			*GWR*		
	Case-I	*Case-II*	*Case-III*	*Case-I*	*Case-II*	*Case-III*	*Case-I*	*Case-II*	*Case-III*
Adj. R^2	0.9941	0.6851	0.9925	0.003	0.0022	0.003	0. 9974	0.7156	0.9920
AICc	641.192	779.363	646.394	81132	151161	84727	618.9038	778.8805	651.0711

The predicted values of GWR for Case -1 (Table 12), FIDs – 18 (MH) with predicted values is 108578.4, similarly FID -17 (MP) has predicted cases (11045.25), FID – 23 (DL) with cases (39150.27), FID – 28 (TN) with 43040.47 and FID- 29 (TS) with cases (6228.23). The following equation y = 0.09130 + -0.00002 x and R2 = 0.05487770288 was generated for case –I.

Table 12: Summary table of GWR predicted model with explanatory variables –Case-I.

FID	*Shape*	*SOURCE_ID*	*Con_ Mar_15*	*Con_Apr_7*	*Con_Apr_12*	*Con_ May_12*	*Con_June_1*	*Predicted*	*Num_ Neighs*
0	Polygon	0	0	10	1	33	33	600.4831	34
1	Polygon	1	0	0	1	1	4	247.8552	34
2	Polygon	2	0	0	28	65	1272	2533.877	34
3	Polygon	3	0	30	48	747	3815	7810.23	34
4	Polygon	4	0	18	14	174	293	1189.009	34
5	Polygon	5	0	9	15	59	498	1370.694	34
6	Polygon	6	0	0	1	1	2	115.0099	34
7	Polygon	7	0	0	1	1	2	114.5943	34
8	Polygon	8	0	7	2	7	70	489.9623	34
9	Polygon	9	0	105	426	8541	16779	22667.46	34
10	Polygon	10	14	49	141	730	2091	3886.695	34
11	Polygon	11	0	6	21	59	331	795.1511	34
12	Polygon	12	2	75	235	879	2446	4717.531	34
13	Polygon	13	0	0	16	160	610	1081.086	34
14	Polygon	14	7	128	172	862	3221	7738.052	34

15	Polygon	15	24	295	193	519	1269	5859.933	34
16	Polygon	16	0	0	0	0	0	97.28564	34
17	Polygon	17	0	104	478	3785	8089	11045.25	34
18	Polygon	18	34	490	1574	23401	67655	108578.4	34
19	Polygon	19	0	2	1	2	71	484.653	34
20	Polygon	20	0	0	0	13	27	288.1443	34
21	Polygon	21	0	1	1	1	1	291.8703	34
22	Polygon	22	0	0	0	0	43	346.6083	34
23	Polygon	23	7	445	1023	7233	19844	39150.27	34
24	Polygon	24	0	5	6	12	70	379.6145	34
25	Polygon	25	1	57	126	1877	2263	3373.274	34
26	Polygon	26	4	200	672	3988	8831	14221.46	34
27	Polygon	27	0	0	0	0	1	238.1447	34
28	Polygon	28	1	411	1020	8002	22333	43040.47	34
29	Polygon	29	3	159	393	1275	2698	6228.23	34
30	Polygon	30	0	0	2	152	313	636.1787	34
31	Polygon	31	13	174	430	3573	7823	13970.76	34
32	Polygon	32	1	16	26	68	907	2361.41	34
33	Polygon	33	0	69	107	2063	5501	10357.58	34
34	Polygon	34	0	5	41	414	1948	3473.054	34
35	Polygon	35	1	161	400	2018	3679	7403.541	34

Daywise increase in states

Figure 10 represents the graphs for 12 different states with MH and TN having maximum cases and the states, namely KL and KA where the curve is flattened. The graphs indicate the day wise increase in the number of confirm cases. Figure 11a represents each curve for all different states in a single graph, whereas Figure 11b shows the day-wise increase in the total cases in India. Figure 12 portraits the samples tested with the confirmed cases in twelve different states namely MH, MP, TN, TS, AP, UP, KA, KL, WB, DL, GJ and RJ.

Tested samples in few states

The graph for number of people tested was obtained on daily update taken from Indian Council of Medical Research (ICMR) website. The total number of samples tested as on June 15, 2020 in India in different states i.e, TN (729002), MH (671348), RJ (609296) and AP (567375) against the confirmed cases in TN (44661), MH (107958), RJ (12694) and AP (6163). It was observed that few states have very less number of testing labs with which the cases are unknown.

Lockdown period graphs in India

The lockdown period (Figure 13) in all India is divided into four phases initially; later fifth phase named as unlock period (Figure 14) is announced. The first phase was from March 25 to April 14, 2020; the second phase was from April 15 to May 3, 2020. During this lockdown period entire country observed no movement outside the home- 'Stay Home Stay Safe' was the quotes referred. The third phase started from May 4 to May 17, 2020 (only 14 days) but the lockdown was extended as the fourth lockdown phase from May 18 to May 31, 2020. After the fourth lockdown, few working sectors and malls were opened with strict guidelines of social distancing and frequently sanitizing to prevent

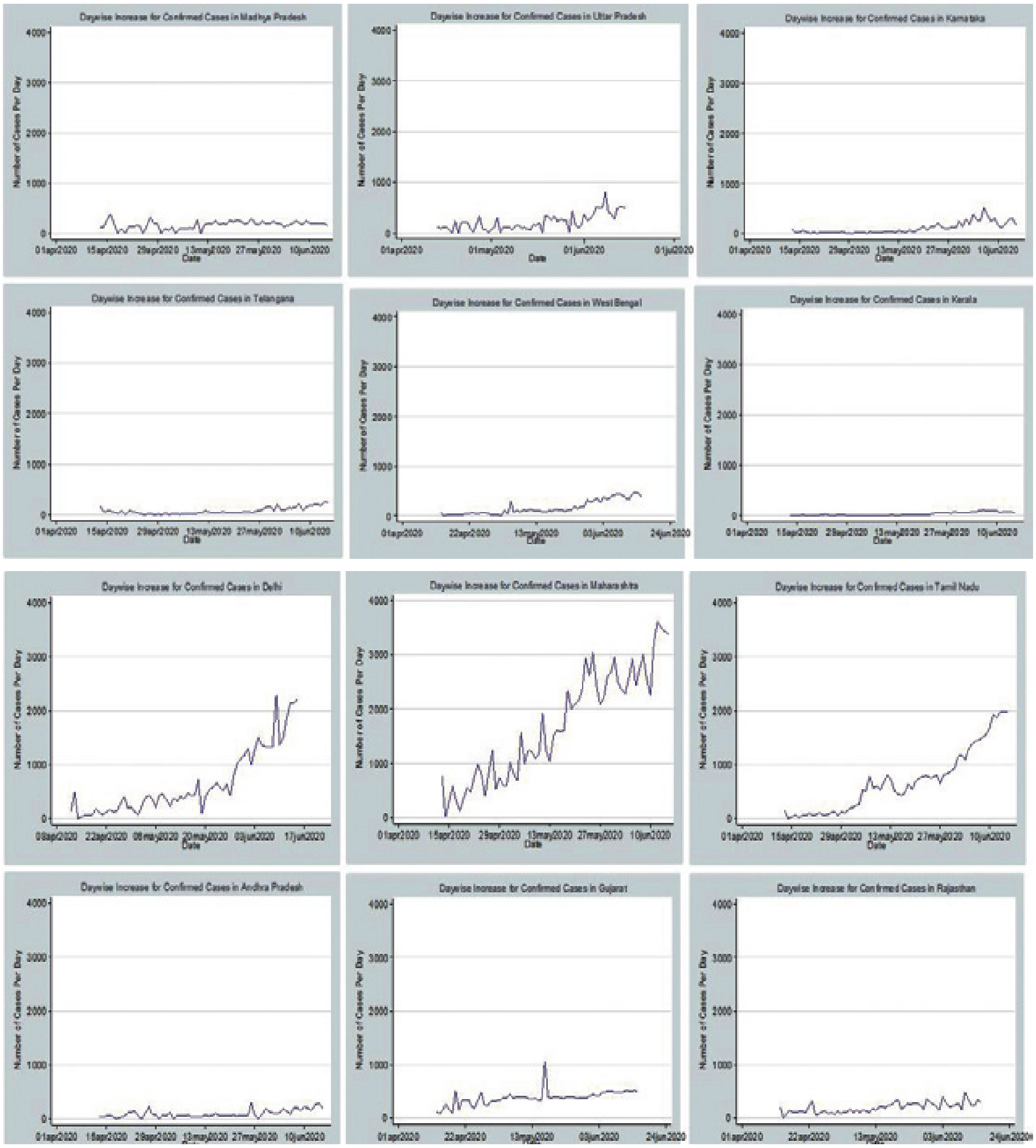

Figure 10: Day wise increase in confirmed cases in 12 different states of India as on June 15, 2020.

COVID-19 attack. From June 1 to June 31, 2020 was stated as fifth lockdown period in the containment zones. Yet, public transport is not in move.

Based on the graphs obtained the slope is calculated as in Table 13. The percentage for each slope is obtained as m1 (-5.714 %), m2 (39.393%), m3 (6.521%) and m4 (46.938%).

Table 13: Slope in each lockdown period.

Slope of lockdown period	
m1	0.035
m2	0.033
m3	0.046
m4	0.049
m5	0.072

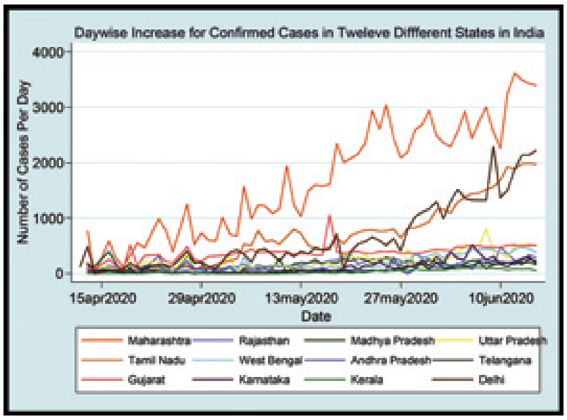

Figure 11a: Day- wise increase of confirmed cases in all 12 states in single graph.

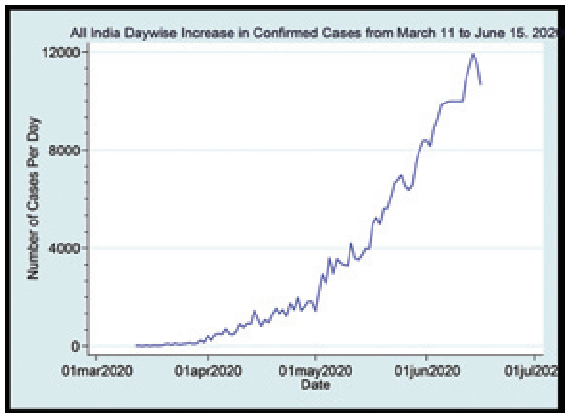

Figure 11b: Day- wise increase of confirmed cases in India.

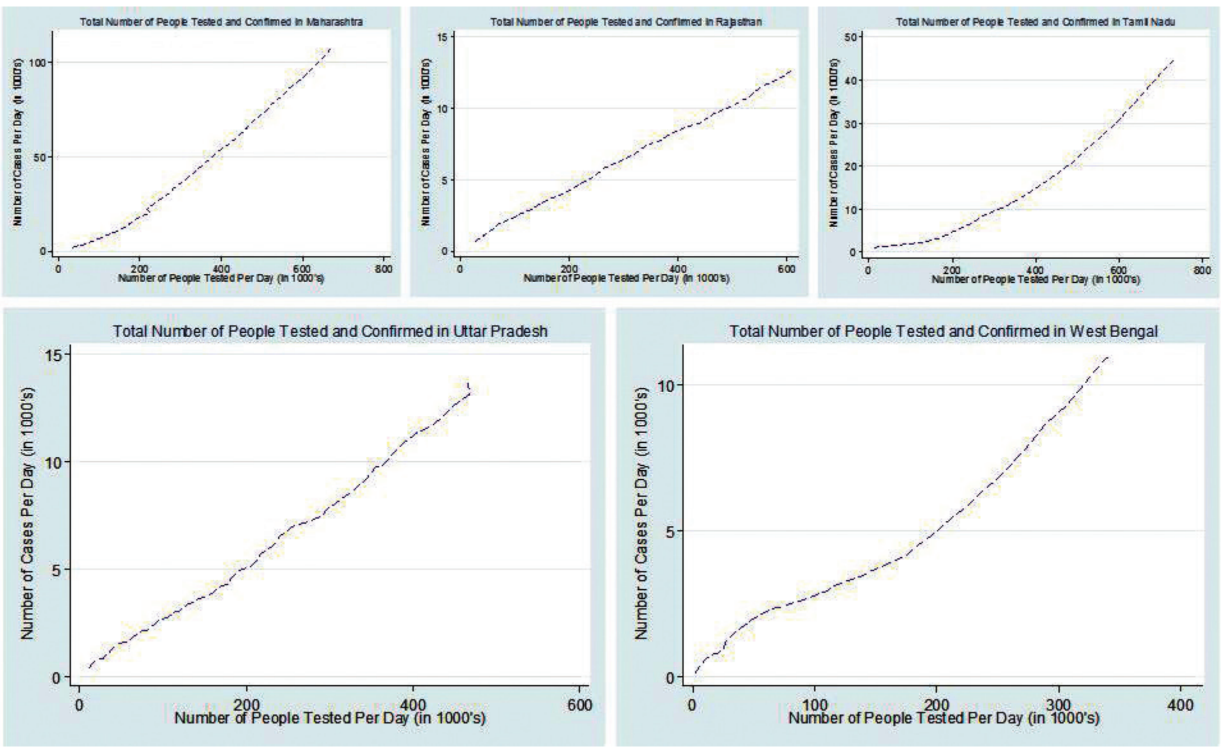

Figure 12: The graph for confirmed cases and samples tested in few states.

Based on the research work carried out by Mollalo et al. on US continental state where GWR and MGWR were proposed for COVID-19 data of US along with SLM, SEM and OLS models. This study extends its models for three different models of OLS, GLR and GWR with three different cases in each model. In this work, the explanatory were confirmed cases of different weeks of COVID-19 whereas Mollalo et al. modelled on explanatory variables of household income, nursing practioners, number of hospitals and soon. The predict values are mentioned in this work along with the day-wise increase and number of people tested in few states. The lockdown graph with the slope is the highlight of this study (Figure 15).

4. Conclusion

GWR model achieved the highest goodness-of-fit among OLS and GLR models, the results of confirmed cases and the findings of the study proved. GWR obtained AICc (618.9038) and Adj-R^2 (0.9974) whereas GLR achieved AICc (81132) and Adj-R^2

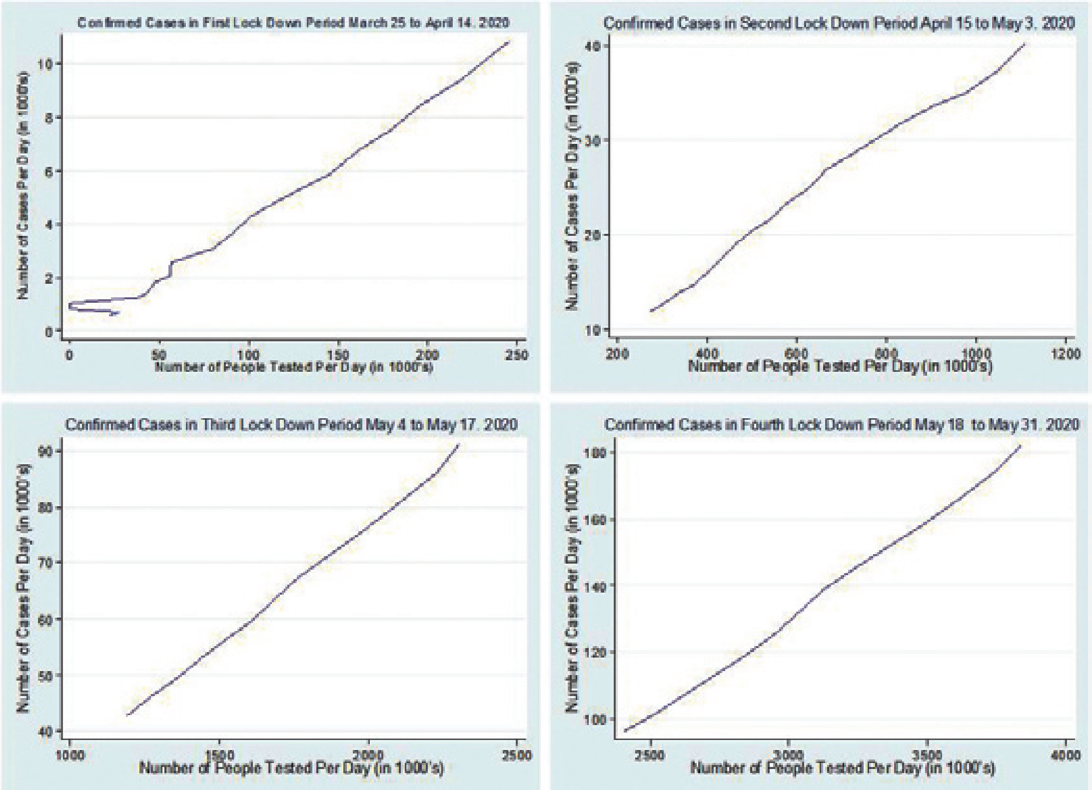

Figure 13: Confirmed cases in four lockdown period in India.

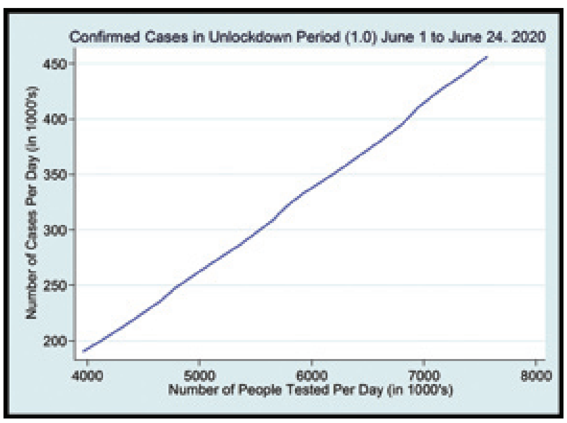

Figure 14: Confirmed cases in unlock down period in India.

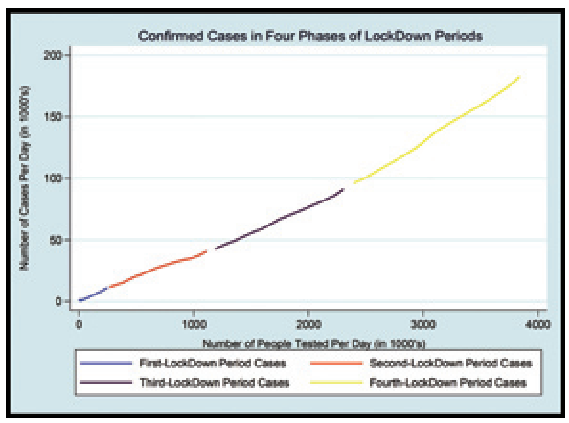

Figure 15: All lockdown period in a single graph.

(0.0034) and OLS produced AICc (641.1929) and Adj-R^2 (0.9941). As stated earlier in discussion since GWR model is a local model compared to GLR and OLS which are global. However, the spatial variability of GWR or OLS or GLR in different countries may reflect different behavior of COVID-19 cases in response to the selected explanatory variables. This study will help in taking the decision for arranging well- equipment testing labs in less availability of testing labs in states of India.

Conflicts of interest

Authors declare no conflicts of interest.

Abbreviations

AICc: Akaike's Information Criterion; DF: Degree of Freedom; SE: Standard Error; VIF: Variance Inflation Factor; OLS: ordinary least models, SLM: spatial lag model; SEM: spatial error model; PLS: Partial Least Square

References

[1] Bashir MF, Ma B, Komal B, Bashir MA, Tan D, et al.Correlation between climate indicators and COVID-19 pandemic in New York, USA. Sci Total Environ. 2020; 728:138835.

[2] WHO. Water, sanitation, hygiene and waste management for the COVID-19 virus. 2020. Available from: https://apps.who.int/iris/handle/10665/331499

[3] Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, et al. First case of 2019 novel coronavirus in the United States. The New England Journal of Medicine. 2020; 382:929–936.

[4] Mollalo A, Vahedi B, Rivera KM. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. Sci Total Environ. 2020; 728:1–8.

[5] Wang C, Zhang J, Yan X, 2012. The Use of Geographically Weighted Regression for the Relationship among Extreme Climate Indices in China. Mathematical Problems in Engineering. Hindawi Publishing Corporation. 2012; 4:1–15.

[6] Calvo E, Escolar M. The Local Voter: A Geographically Weighted Approach to Ecological Inference. American Journal of Political Science. Wiley. 2003; 47(1):189–204.

[7] Comber A, Chi K, Huy MQ, Nguyen Q, Lu B, et al. Distance metric choice can both reduce and induce collinearity in geographically weighted regression. Environment and Planning B: Urban Analytics and City Science. Sage Journal. 2018; 47(3):489–507.

[8] Franch-Pardo I, Napoletano BM, Rosete-Verges F, Billa L. Spatial analysis and GIS in the study of COVID-19. A review. Sci Total Environ. 2020; 739:140033.

[9] Gupta A, Banerjee S, Das S. Significance of geographical factors to the COVID-19 outbreak in India. Model Earth Syst Environ. 2020; 1–9.

[10] Boulos MNK, Geraghty EM. Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. Int J Health Geogr. 2020 Mar 11;19(1):8.

[11] Krishnakumar B, Rana S. COVID 19 in INDIA: Strategies to combat from combination threat of life and livelihood. J Microbiol Immunol Infect. 2020; 53(3):389–391.

[12] Pulla P. Covid-19: India imposes lockdown for 21 days and cases rise. BMJ. 2020; 368:m1251.

[13] Ward MD, Gleditsch KS. Spatial regression models. Sage Publications. 2018; 155:1–128. ISBN: 9781544328836.

[14] Biuand RS. Regression modeling with spatial dependence: an application of some class selection and estimation methods. Geographical Analysis. 1984; 16(1):25–37.

[15] Srinivasan S. Spatial regression models. Encyclopedia of GIS, Springer International Publishing Switzerland, 2016; 1–6.

[16] Anselin L, Arribas-Bel D. Spatial fixed effects and spatial dependence in a single cross-section. Papers in Regional Science. 2013; 92(1):3–17.

[17] Mogaji KA, Lim HS. A GIS-based linear regression modeling approach to assess the impact of geologic rock types on groundwater recharge and its hydrological implication. Model Earth Syst Environ. 2019; 6:183–199.

[18] Oshan TM, Smith JP, Fotheringham AS. Targeting the spatial context of obesity determinants via multiscale geographically weighted regression. Int J Health Geogr. 2020; 19(1):1–17.