International Journal of Community Medicine and Public Health
Sharma M et al. Int J Community Med Public Health. 2026 May;13(5):2282-2289
http://www.ijcmph.com
pISSN 2394-6032 | eISSN 2394-6040

DOI: https://dx.doi.org/10.18203/2394-6040.ijcmph20261410

Original Research Article

Enhancing COVID-19 analysis using adaptive robust geographically weighted regression: a global perspective

Megha Sharma*, Shalini Chandra

Department of Mathematics and statistics, Banasthali Vidyapith, Rajasthan, India

Received: 16 January 2026
Revised: 06 April 2026
Accepted: 17 April 2026

***Correspondence:**
Dr. Megha Sharma,
E-mail: meghasharma@banasthali.in

ABSTRACT

Background: Understanding the global spread and impact of COVID-19 requires analytical approaches that capture spatial heterogeneity and data irregularities. Traditional regression methods often fail to address issues such as outliers and heteroscedasticity, limiting their effectiveness in modeling pandemic data.
Methods: This was an ecological and spatial analytical study using secondary global COVID-19 data. This study employs an adaptive robust geographically weighted regression (AR-GWR) model integrated with the γ-divergence technique to enhance robustness against outliers and non-constant variance. Unlike classical regression and standard geographically weighted regression (GWR), the AR-GWR model incorporates adaptive bandwidth selection, enabling automatic optimization of spatial smoothness and robustness parameters. This improves localized parameter estimation and predictive performance.
Results: The findings reveal substantial spatial variation in COVID-19 outcomes across countries. Nations with advanced healthcare systems, such as South Korea and France, report higher case numbers despite high human development index (HDI) scores. In contrast, Sub-Saharan African countries exhibit relatively lower case and mortality rates, potentially due to demographic and geographic factors. The AR-GWR model identifies healthcare infrastructure and preexisting health conditions as significant determinants of COVID-19 mortality, while demographic factors primarily influence infection rates. Compared to traditional models, AR-GWR demonstrates superior predictive accuracy and better handling of spatial non-stationarity.
Conclusions: The study highlighted the effectiveness of advanced spatial regression techniques, particularly AR-GWR with adaptive bandwidth selection, in modelling complex pandemic data. By accounting for spatial heterogeneity and data irregularities, the model provides more reliable insights into global health patterns. These findings can support improved pandemic preparedness and informed public health decision-making.

Keywords: AR-GWR, Bandwidth selection, COVID-19, γ-divergence, GWR, OLS, Predictive modelling, Spatial regression

INTRODUCTION

The COVID-19 pandemic has underscored the importance of statistical modeling in understanding disease transmission patterns. While the immediate crisis has passed, researchers continue to analyze COVID-19 data to uncover long-term trends and refine predictive methodologies.[1,2] Classical regression models were initially used to assess global transmission patterns. However, these models assume spatial homogeneity, overlooking the local variations in disease spread. Spatial regression techniques, particularly geographically weighted regression (GWR), have been widely used to capture location-specific differences and study the spatial heterogeneity of COVID-19 through localized parameter estimation. By accounting for regional variations in

transmission dynamics, it provides a detailed understanding of the pandemic's spread.[3,4] However, GWR is highly sensitive to violations of assumptions, such as the presence of outliers, heteroscedasticity, and multicollinearity, which can distort estimates and reduce model reliability. These challenges highlight the need for an improved approach that retains the advantages of GWR while addressing its limitations. To address multicollinearity, the selection of predictor variables must account for collinearity, while robust extensions of GWR should be employed to enhance the model's resilience to outlier influence and heteroscedasticity, thereby improving the reliability of spatial inference.[5]

A robust extension of GWR has been developed through two primary approaches.[5] The first involves the exclusion of observations with excessively large residuals prior to model fitting, though this method may lead to the loss of critical information, making it unsuitable for pandemic-related data. The second approach employs an iterative fitting process that assigns reduced weights to observations with large residuals, thereby mitigating their influence on parameter estimation. This down-weighting technique has been further extended through various methodologies.[6] For instance, a Bayesian GWR model incorporating non-constant variance priors has been introduced to regularize outliers.[6] Alternative robust estimation techniques have also been explored, including least absolute deviation estimation, M-quantile estimation, and asymmetric absolute loss-based geographically weighted quantile regression.[7-9]

These approaches primarily focus on the robust estimation of spatially varying parameters while placing less emphasis on the selection of the bandwidth parameter, which controls the extent of spatial smoothing and significantly influences model performance and estimation accuracy. Although these methods incorporate tuning parameters to enhance robustness against outliers, their selection lacks a systematic framework and is often based on rule of thumb rather than rigorous optimization.[10] To address these limitations, the adaptive robust geographically weighted regression (AR-GWR) approach has been introduced for robust estimation and inference in GWR models.[10] AR-GWR utilizes γ-divergence, a general class of robust objective functions that encompasses the log-likelihood function as a special case. The robust objective function in AR-GWR can be efficiently optimized using the majorization-minimization (MM) algorithm.[11] This method allows AR-GWR to automatically select the best bandwidth for the model using a reliable cross-validation process, and it also provides more accurate estimates of the uncertainty in the model's coefficients, even in the presence of outliers.

Building on the robust methodologies discussed earlier, this study applies the AR-GWR framework to a comprehensive COVID-19 dataset. The goal is to analyze spatial disparities in disease spread and identify key risk factors while effectively addressing outliers and heteroscedasticity in the data. In addition to applying AR-GWR, this study also compares its performance with OLS regression and GWR models using diagnostic measures such as Adjusted R^2, RMSE, and AIC, BIC to highlight the advantages of AR-GWR in handling data irregularities and improving model accuracy. By utilizing AR-GWR, this analysis focuses on COVID-19 transmission at the global level, with countries serving as the spatial units of analysis. This approach facilitates the exploration of pandemic trends across different regions, offering valuable insights into the global distribution of COVID-19 and emphasizing regional variations influenced by factors such as population density, healthcare accessibility, comorbidities, environmental conditions, and socioeconomic factors.

METHODS

This study was an ecological and spatial analytical investigation based on secondary data on COVID-19 collected at the global level. It examined country-level variations to understand spatial patterns and relationships in pandemic outcomes across different regions. In this section, detailed information regarding the data sources, variables considered, and the statistical models employed

Data description and preparation

This study utilized data from 191 countries, excluding North Korea, Turkmenistan, Liechtenstein, and Vatican City due to the unavailability of COVID-19 data. The dataset comprises cumulative COVID-19 cases and deaths per million population (CCPM, CDPM) across countries up to May 2024, providing a standardized basis for comparison and analysis of spatial variations in disease spread. To ensure geographic accuracy, administrative boundary shapefiles were sourced from the World Bank Spatial Boundaries Database and integrated using ArcGIS Desktop 10.7 and then further analysis done in R-software.

On the basis of existing literature, this study selects 17 independent variables, encompassing demographics, healthcare infrastructure, socioeconomic conditions, and environmental factors. These include the prevalence of cancer (PC), prevalence of cardiovascular disease (PCD), prevalence of hypertension (PH), prevalence of diabetes (PD), death rate from respiratory infections (DRRI), death rate due to air pollution (DRAP), human development index (HDI), current healthcare expenditure (CHE), hospital beds per thousand (HBPT), improved water facility access (IWF), improved sanitation facilities access (ISF), tobacco usage (TU), older population (AGE65+), population density (PDN), urban population share (UP), life expectancy (LE), proportion of PM 2.5 (PM). These variables are expected to significantly influence the transmission dynamics of COVID-19.[13] All variables were sourced from Our World in Data (ourworldindata.org), a comprehensive repository of global research and datasets. To maintain data integrity,

extensive preprocessing was performed, including handling missing values and duplicates through cross-referencing multiple sources. The multi- source data was standardized and normalized to a common scale to facilitate meaningful comparisons. The spatial integration of variables enables the examination of geographic variations and their potential impact on the pandemic's spread.

Data visualization and exploration

This study utilized choropleth maps, LISA maps, and boxplots to examine the spatial distribution and trends of CCPM and CDPM. Choropleth maps depict regional disparities, LISA maps highlight spatial clusters and outliers, and boxplots provide insights into the distribution across regions. Figures 1 to 3 illustrate these visualizations, offering a detailed representation of CCPM and CDPM across various countries. Specifically, Figures 1(A) and 1(B) present choropleth maps with classified regions, where colour gradients represent variations in intensity.

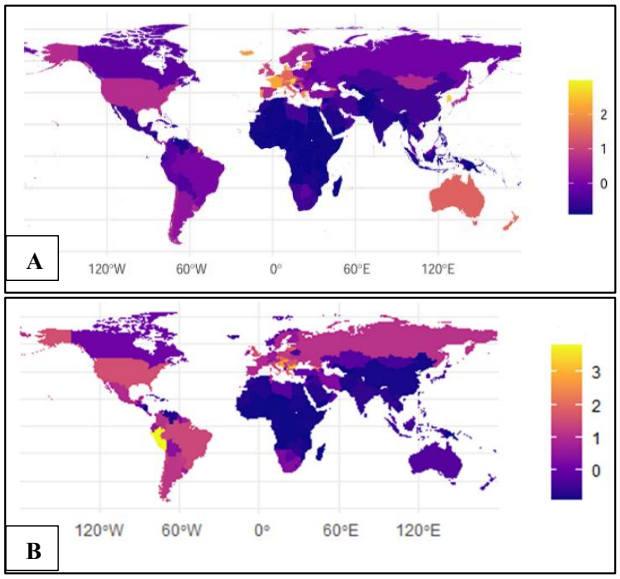

Figure 1: Choropleth maps for: A) CCPM; B) CDPM in world countries.

As shown in Figure 1, European and Asian countries such as San Marino, Cyprus, Brunei, Austria, South Korea, and France exhibit high COVID-19 cases per million population, despite their advanced healthcare infrastructure and elevated HDI. In contrast, Sub-Saharan African nations such as Yemen, Niger, Chad, Tanzania, and Nigeria report lower case and death rates, potentially attributable to lower population densities, larger land areas, and younger populations experiencing milder infections. Although countries such as the USA, India, Brazil, Russia, and the UK report high absolute numbers of cases and deaths, their per capita rates place them among moderately affected nations globally. The Moran's I statistic indicates significant spatial clustering in both COVID-19 cases and deaths (0.63 and 0.67, p value =0.0001), confirming a non-random distribution pattern. However, as Moran's I cannot detect localized clusters, LISA was applied to further investigate spatial structure. Figure 2(A, B) highlights high-case clusters in eastern and western Europe, with additional mortality clusters in Latin America. Regions with low cases and deaths, such as Eastern Asia, Southern Asia, and Sub-Saharan Africa, also form distinct clusters, reinforcing the presence of strong spatial dependence in COVID-19 transmission patterns across countries.

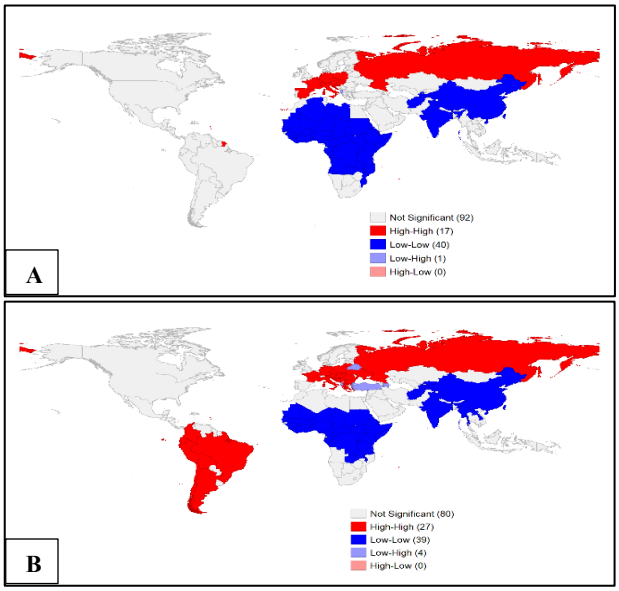

Figure 2: LISA maps for: A) CCPM; B) CDPM in world countries.

Figure 3 presents a boxplot visualization of the distribution of CCPM and CDPM. Both variables are the right skewed indicated concentration of lower values with a few extreme observations.

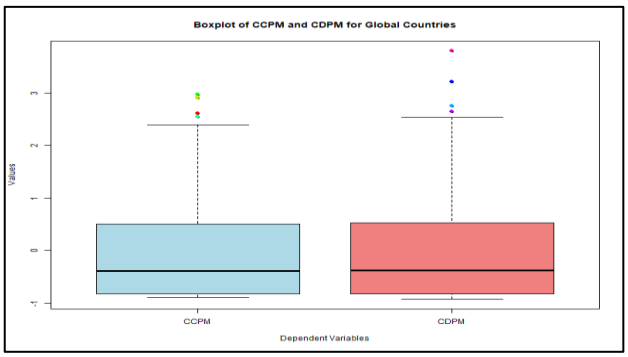

Figure 3: Boxplot visualization for the A) CCPM and B) CDPM

For CCPM, five outliers-Austria, Brunei, Cyprus, San Marino, and South Korea- demonstrate significantly higher cases counts. Similarly, the distribution of deaths four outliers- Bosnia and Herzegovina, Bulgaria,

Hungary, and Peru- displaying exceptionally high death rates. These outliers suggest a disproportionately severe impact in these regions, warranting further investigation into the underlying factors contributing.

The models and estimators

To estimate the linear relationship between a dependent variable and one or more independent variables, the classical linear regression model (CLRM) was employed. The model is specified as

$$y = X\beta + \in \qquad (1)$$

where, y is the $n \times 1$ vector of the observations on dependent variable, X is an $n \times m$ matrix of independent variables, $\in$ is the $n \times 1$ vector of the random error. β is the $n \times 1$ vector of the parameters which is estimated by the minimizing sum of the residuals square and obtained as

$$\hat{\beta} = (X^T X)^{-1}\ X^T y\,. \qquad (2)$$

The estimator in eq. (2) is called ordinary least squares estimator (OLS). Under a set of standard assumptions- linearity in parameters, full column rank of X, exogeneity, homoscedasticity, no autocorrelation, and normality of the error terms- the OLS estimator is the best linear unbiased estimator, as stated by the Gauss-Markov theorem. However, the presence of spatial dependence, wherein observations are correlated across geographic units, violates the CLRM assumption of error term independence. As a consequence, the OLS estimator may become inefficient, and the resulting statistical inference unreliable. To address this issue, spatial regression models such as the spatial lag model (SLM), the spatial error model (SEM), and geographically weighted regression (GWR) are employed to account for spatial dependence by incorporating the spatial structure among observations. While SLM and SEM model global spatial effects through a spatial weight matrix, GWR provides a more flexible, locally adaptive approach by allowing regression coefficients to vary across geographic locations.

The GWR models estimated local interactions between the dependent variables and independent variables by fitting a regression model to each feature (spatial unit) in the dataset. The GWR model for each feature is:

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

where y_i represents dependent variable at a specific location i, β_{io} stands for the intercept parameter at that same location i, β_{ij} symbolizes the local regression coefficient pertaining to the j^{th} explanatory variable at location i, X_{ij} signifies the value of the j^{th} explanatory variable at location i, and ϵ_i corresponds to the random error observed at location i. The given parameters estimate for each independent variable at i^{th} location is given by

$$\hat{\beta}(i) = (X^T W(i) X)^{-1}\ X^T W(i) y. \qquad (4)$$

The estimator given in eq. (4) is weighted least squares estimator where $\hat{\beta}(i)$ is $m \times 1$ vector of parameter estimates, $W(i)$ is spatial weight matrix calculated by the Gaussian kernel function which is defined as

$$W_k(i) = \begin{cases} e^{-\left(\frac{d_{ik}}{bw}\right)^2}, & if\ k \in \{N_i\} \\ 0, & if\ k \notin \{N_i\} \end{cases} \qquad (5)$$

where d_{ik} is the distance between features location i and k with bandwidth bw derived from the Euclidean distance between observation location and neighbouring points, this measure ensures that the region remains influenced by proximate neighbours within this radius. The set N_i includes observations within this N^{th} nearest neighbour distance. Weights are zero for observations beyond this range. Kernel function assigns higher weights to observations that are closer to the calibration location i. To fit GWR model, the kernel bandwidth is estimated through cross-validation (CV) using all feature locations, followed by weight calculation using (5). CV function is outlined as

$$CV(bw) = \sum_{i=1}^{n} [y_i - \hat{y}_{\neq i}(bw)]^2 \qquad (6)$$

where, $\hat{y}_{\neq i}(bw)$ is the estimated value of y achieved by the excluding the data point at the i^{th} location during prediction. The bandwidth bw will be derived through an iterative process aiming to minimizing the CV score.[5]

GWR model effectively accounts for spatial heterogeneity; however, its sensitivity to outliers and heteroscedasticity limits its robustness. The Adaptively Robust GWR (AR-GWR) model, integrates a γ-divergence-based objective function to mitigate the influence of extreme values, enhancing the model's stability and reliability in spatial analysis.[11] The AR-GWR model follows a structure similar to the GWR model, as shown in eq (3). However, instead of using conventional weighted least squares for parameter estimation, AR-GWR varying parameters to be estimated through geographically weighted likelihood, as given by

$$\hat{\beta}(i) = argmax_{\beta\sigma^2} L_i[\beta, \sigma^2]\,, \qquad where$$

$$L(\beta, \sigma^2) = \sum_{k=1}^{n} w\left(|dik|/_{bw}\right).\log\ \phi[y_i; X\beta,, \sigma^2] \quad (7)$$

In eq (7), $w(\cdot)$ is a spatial weight function. Further, adding robustness by replacing the standard likelihood with γ-divergence function, given by

$$D_i(\beta,\sigma^2) = \frac{1}{\gamma}\log\left(\sum_{k=1}^{n} w\left(\frac{|dik|}{bw}\right).\phi\left[y_i; X^T\beta,\sigma^2\right]^\gamma\right) + \frac{\gamma}{2(1+\gamma)}log\sigma^2 \quad (8)$$

Where γ and bw is two tunning parameters controlling the robustness of the resulting estimators, which selected by the robust cross validation (RCV) criterion given by

$$\mathrm{RCV}(bw;\gamma) = \frac{1}{\gamma}\log\left(\sum_{\substack{k=1\\k\neq i}}^{n}\phi\left[y_i; X^T\hat{\beta},\hat{\sigma}^2\right]^\gamma\right) + \frac{\gamma}{2(1+\gamma)}log\left(\sum_{\substack{k=1\\k\neq i}}^{n}\hat{\sigma}^2\right) \quad (9)$$

RESULTS

Model implementation and bandwidth selection

The CLRM results indicate that independent variables significantly influence CCPM and CDPM, with adjusted R^2 values of 0.61 and 0.55, respectively, suggesting moderate explanatory power. However, diagnostic tests reveal multicollinearity is not a concern, as evidenced by variance inflation factors (VIFs) remaining below the conventional threshold of 10. However, the Durbin-Watson statistics (CCPM: 2.89, CDPM: 2.74, p value >0.05) suggest residual autocorrelation, while Moran's I test applied to the residuals (CCPM: 0.32, CDPM: 0.28, p value <0.05) confirms statistically significant spatial autocorrelation. These findings highlight the violation of independence assumptions and the inadequacy of the OLS framework in the presence of spatial dependence.

These findings underscore the violation of key independence assumptions and the limitations of the OLS framework in capturing spatial dependence. To address this dependency, the study employs GWR, which generates localized parameter estimates using an adaptive Gaussian kernel, thereby accounting for spatial heterogeneity. However, residual diagnostics reveal persistent heteroscedasticity, as indicated by the Breusch-Pagan test (CCPM: 37.17, CDPM: 24.572, p value <0.05), along with the presence of outliers evident in Figure 4. These diagnostic issues necessitate the use of the AR-GWR model, which enhances robustness against these issues. The optimal bandwidth selection reflects this adjustment, with AR-GWR selecting slightly higher bandwidths (CCPM: 0.186 versus 0.156 for GWR; CDPM: 0.214 versus 0.190 for GWR), enabling better local adaptation. Furthermore, AR-GWR incorporates a gamma-divergence tuning parameter (0.054 for CCPM, 0.027 for CDPM), improving model stability and accuracy in capturing the spatial dynamics of COVID-19.

Model diagnostics and robustness comparisons

To evaluate model performance, we compare key diagnostic measures, including adjusted R^2, RMSE, AIC and BIC, as presented in Table 1. Among all models, the Adaptive AR-GWR shows the best performance, achieving the highest adjusted R^2 and the lowest RMSE, indicating its strong ability to capture spatial heterogeneity and manage extreme values effectively. The lower AIC and BIC values further confirm that AR-GWR offers a more reliable and efficient approach for analyzing COVID-19 spatial patterns.

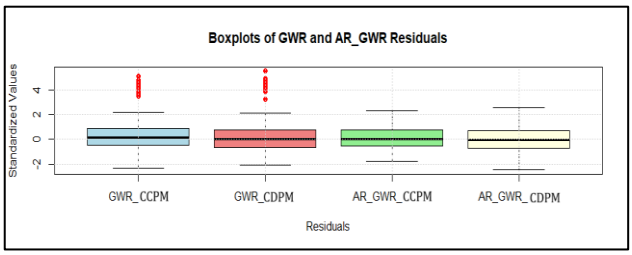

Figure 4: Boxplot visualization for the residual of GWR and AR-GWR model.

Table 1: Summary of measures of model's diagnostics measures.

Models		Adjusted R^2	RMSE	AIC	BIC
CCPM	CLRM	0.613	0.599	369.87	421.911
	GWR	0.7323	0.4863	342.121	416.135
	AR-GWR	0.7612	0.4576	339.69	412.49
CDPM	CLRM	0.5598	0.421	369.875	421.911
	GWR	0.712	0.354	299.85	369.56
	AR-GWR	0.7534	0.334	295.98	367.54

To further assess the impact of outliers, we examine the residuals of GWR and AR-GWR models. GWR exhibits heavy-tailed residuals with several extreme values, indicating sensitivity to outliers. In contrast, AR-GWR yields a more symmetric distribution with reduced influence from extremes, highlighting its ability to mitigate heteroscedasticity. As shown in Figure 4, AR-GWR demonstrates greater residual stability.

This robustness is further confirmed by the Breusch-Pagan test results (CCPM: 17.12, CDPM: 14.342, p value >0.05), indicating homoscedastic errors and minimal outlier effects. Overall, while GWR improves spatial

modeling over OLS, it remains susceptible to data irregularities, whereas AR-GWR effectively addresses these limitations, offering a more reliable framework for understanding COVID-19 transmission dynamics across spatial regions. Since AR-GWR outperforms other competing methods, we summarize its key findings in the following section.

AR-GWR model summary

AR-GWR model generates spatially variable regression coefficients that vary across locations. Each spatial unit (country) is associated with distinct regression coefficients and potentially different sets of statistically significant variables. However, to provide an overview of the average effects of all the selected variables on CCPM and CDPM across global countries, the summary of coefficient estimates is presented in Table 2.

The results presented here provide important insights into the spatial dynamics and determinants of COVID-19 across countries. The observed differences in model performance, along with the identification of significant variables, highlight the complexity and spatial heterogeneity inherent in pandemic data. These findings form the basis for further interpretation and are discussed in the following section in the context of existing literature and methodological considerations.

Table 2: Summary table for the AR-GWR model for the CCPM and CDPM.

Variable	CCPM				CDPM			
	Coeff. Est.	Std. Error	t-value	Pr (>|t|)	Coeff Est.	Std. error	t-value	Pr (>|t|)
Intercept	-0.04	0.051	-0.78	0.458	-0.05	0.048	-1.00	0.284
PC	0.141	0.06	6.43	0.073	0.164	0.058	2.83	0.005
PCD	0.000	0.041	0.00	0.999	0.193	0.045	4.24	2.1e-5
PH	0.130	0.04	0.76	0.050	0.142	0.039	3.57	0.000
PD	0.127	0.030	0.88	0.075	0.158	0.034	3.58	0.000
DRRI	0.029	0.046	0.64	0.515	0.209	0.050	4.16	3.4e-5
AGE65+	0.07	0.095	0.82	0.414	0.228	0.078	5.38	7.4e-7
PDN	0.059	0.028	2.07	0.039	-0.03	0.026	-1.17	0.243
UP	0.17	0.040	4.39	0.000	0.128	0.042	2.99	0.003
LE	0.43	0.067	6.43	0.000	0.368	0.062	5.90	2.8e-8
TU	0.11	0.031	3.71	0.000	0.095	0.037	0.953	0.349
ISF	0.17	0.042	4.02	0.000	0.038	0.048	2.84	0.054
IWF	-0.03	0.036	-0.97	0.340	-0.02	0.032	-0.76	0.447
DRAP	0.01	0.035	0.51	0.607	0.152	0.043	3.47	0.0006
PM	0.19	0.048	0.19	0.084	0.209	0.053	3.88	0.000
HDI	0.11	0.044	2.50	0.013	0.079	0.040	1.97	0.048
CHE	0.10	0.039	3.35	0.000	-0.03	0.098	-0.35	0.728
HBPT	-0.11	0.062	-1.81	0.070	-0.06	0.052	-0.34	0.691

DISCUSSION

The present study emphasized the importance of incorporating spatial heterogeneity and robustness in modelling global COVID-19 dynamics. The findings reveal clear limitations of classical regression approaches and demonstrate the advantages of advanced spatial models, particularly the AR-GWR framework.

The results from CLRM indicate moderate explanatory power; however, the presence of significant spatial autocorrelation confirms the violation of independence assumptions. This limitation is consistent with previous studies, which reported strong spatial clustering in COVID-19 cases and highlighted the inadequacy of non-spatial models in capturing geographic variation.[3,4] Similarly, earlier predictive studies relied on linear and piecewise regression techniques which, although useful for trend analysis, do not adequately account for spatial dependence and heterogeneity.[1,2]

The application of the GWR model improved model performance by capturing spatial non-stationarity, which aligns with foundational studies that established GWR as an effective tool for modelling spatially varying relationships.[5,6] The improved fit observed in this study is also consistent with prior COVID-19 spatial analyses. However, similar to previous findings, standard GWR remains sensitive to heteroscedasticity and outliers, which limits its reliability when dealing with noisy and irregular global datasets.[7]

To overcome these limitations, the AR-GWR model was employed, incorporating robustness through γ-divergence and adaptive bandwidth selection. The superior performance of AR-GWR, reflected in improved diagnostic measures and stable residual behavior, is

consistent with earlier studies[10] that demonstrated the effectiveness of adaptive robust approaches in handling outliers and heteroscedasticity. Additionally, this robustness framework is supported by prior methodological advancements, which emphasized the importance of alternative loss functions in spatial modelling.[9]

From a substantive perspective, the findings reveal that COVID-19 transmission and mortality are influenced by a complex interplay of demographic, healthcare, environmental, and socio-economic factors. The positive association of urbanization and life expectancy with case counts is consistent with global observations and supported by earlier studies, which highlight variability in predictive factors across regions.[12]

The association between pre-existing health conditions (such as cancer, cardiovascular diseases, and hypertension) and COVID-19 mortality supports existing clinical and epidemiological evidence indicating that comorbidities significantly increase the risk of severe outcomes. Furthermore, the significant role of environmental factors, particularly air pollution (PM2.5), aligns with previous research linking poor air quality to higher COVID-19 mortality.

The negative relationship between healthcare capacity indicators and mortality further emphasizes the importance of strong healthcare systems in mitigating pandemic impacts. This finding reinforces the role of healthcare preparedness and resource availability in reducing fatal outcomes. However, the lack of significant effects for certain variables, such as sanitation and water access, suggests that their influence may be indirect or context-specific, as also observed in broader global analyses.

Overall, the study confirmed that COVID-19 dynamics are inherently spatial and complex, requiring advanced modelling approaches. The superior performance of AR-GWR demonstrates its effectiveness in addressing both spatial dependence and data irregularities, making it a valuable tool for global health analysis. These findings contribute to the growing body of literature advocating for robust spatial methodologies and provide important insights for policymakers to design targeted, region-specific interventions.

CONCLUSION

This study revealed considerable spatial variation in COVID-19 cases and deaths across countries, influenced by factors such as healthcare infrastructure, population density, mobility, and government responses. While high-HDI nations like South Korea and France report elevated cases per capita, Sub-Saharan African countries show lower numbers, likely due to younger populations and lower population density- though underreporting may play a role. Spatial analysis confirms non-random clustering of cases and deaths, with distinct regional patterns identified through Moran's I and LISA maps. To capture these variations, the study applies the AR-GWR model with γ-divergence, which handles outliers and heteroscedasticity better than traditional methods. This model's adaptive bandwidth improves accuracy, reduces RMSE, and uncovers localized relationships, revealing that strong healthcare systems reduce mortality but may increase reported cases due to better detection. Ultimately, the findings emphasize the value of robust spatial models like AR-GWR for understanding regional dynamics and informing more effective, tailored pandemic responses.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: The study was approved by the Institutional Ethics Committee

REFERENCES

1. Chaurasia R, Singh S, Singh V. Progression of COVID-19 in India: a linear regression analysis. medRxiv. 2021.
2. Kumar A, Awasthi A. Piecewise linear regression for predicting COVID-19 cases in India. SpringerLink. 2020.
3. Kang D, Choi H, Kim JH, Choi J. Spatial epidemic dynamics of the COVID-19 outbreak in China. Int J Infect Dis. 2020;94:96-102.
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;728:138884.
5. Fotheringham AS, Brunsdon C, Charlton M. Geographically weighted regression: The analysis of spatially varying relationships. Wiley; 2002.
6. LeSage JP. A family of geographically weighted regression models. In: Anselin L, Florax RJM, Rey SJ, eds. Advances in spatial econometrics. 1st edn. Springer Berlin, Heidelberg; 2004:241-264.
7. Zhang J, Mei CL. Local least absolute deviation estimation of spatially varying coefficients models: Robust geographically weighted regression approaches. Int J Geogr Inf Sci. 2011;25(9):1417-38.
8. Salvati N, Pratesi M, Giusti C. The use of M-quantile models as an alternative to random effect models in small area estimation. J R Stat Soc Ser A. 2012;175(1):267-86.
9. Chen Y, Okada N, Zhou Y. Geographically weighted regression based on the asymmetric loss function. Stoch Environ Res Risk Assess. 2012;26(6):799-810.
10. Sugasawa S, Murakami D. Adaptively robust geographically weighted regression. J Spat Sci. 2021;66(2):285-98.
11. Hunter DR, Lange K. A tutorial on MM algorithms. Am Stat. 2004;58(1):30-7.

12. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of COVID-19: systematic review and critical appraisal. BMJ. 2020;369.

Cite this article as: Sharma M, Chandra S. Enhancing COVID-19 analysis using adaptive robust geographically weighted regression: a global perspective. Int J Community Med Public Health 2026;13:2282-9.