The Spatial Predicting of COVID-19 Incidence and Its Mortality Based On OLS and GWR Models in Iran

Abolfazl Ghanbari (danizly1358@yahoo.com)
Tehran University: University of Tehran https://orcid.org/0000-0002-0145-0279

Behzad Baradaran
Tabriz University of Medical Sciences

Hamed Ahmadi
Tehran University: University of Tehran https://orcid.org/0000-0002-4690-4043

Maryam Ahmadi
Tabriz University: University of Tabriz

Research

Keywords: COVID-19, Spatial predicting, OLS, GWR, ArcGIS, Iran

Posted Date: March 18th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-297401/v1

Abstract

Background: Within six months of the COVID-19 outbreak, 350279 people were infected, and 20125 people died of COVID-19 in Iran. There is an urgent need to find the most accurate effective indicators on this disease's outbreak in order to control and predict.

Methods: We examined the effect of 36 demographic, economic, environmental, health infrastructure, social, and topographic independent variables on the COVID-19 infection and mortality rates using the ordinary least squares (OLS) model in ArcGIS 10.5. Regarding adjusted R-squared>0/7, we selected 20 variables for COVID-19 infection rate and 16 variables for the mortality rate. The collinearity problem between the selected variables resolved after using the variance inflation factor (VIF). Then, we performed the OLS and geographically weighted regression (GWR) models in ArcGIS 10.5.

Results: Having a large number of men, having a large population, lack of specialist doctors, lack of hospital, having a large urban population, having a large number of people aged 65 and over or older individuals, and high natural mortality rate had the most prominent impact on the COVID-19 infection increasing rate. Also, lack of ICU beds, low number of insured people, lack of subspecialist physicians, and lack of hospital beds had the most prominent impact on increasing of COVID-19 mortality. Then the variables with VIF above 7.5 were removed and finally, high incoming immigrants rate and lack of nurses were identified as two independent variables to predict COVID-19 infection rate. In addition, high incoming immigrants rate and high number of doctor consultation were recognized as two variables to predict mortality rate due to COVID-19. The results of the Akaike information criterion (AIC) and adj.R2 showed that both models were appropriate for these analyses.

Conclusions: Based on our results, there would be a considerable increase in COVID-19 infection in Kerman, Esfahan, and Kermanshah provinces. In addition, there would be a remarkable decrease in COVID-19 infection in Khuzestan, Lorestan, Azarbayjan Shargi, and Tehran provinces. Regarding COVID-19 mortality, there would be a substantial rise in Fars and Khorasan Razavi provinces. Moreover, our analyses predicted a considerable diminish in COVID-19 mortality in Tehran, Ardebil, Zanjan, Gilan, Golestan, Lorestan, Khuzestan, Bushehr, and Hormozgan provinces.

1. Introduction

The first pneumonia cases of unknown origin were identified in Wuhan, the capital city of Hubei province of China in early December 2019. Further, the pathogen was identified as a novel enveloped RNA beta coronavirus, and was named Severe Acute Respiratory Syndrome Corona Virus 2 (SARS-CoV-2), which has a phylogenetic similarity to SARS-CoV (Manoj and et al., 2020). Then, it rapidly affected many individuals in nearby areas, and next in the whole world. Following that, the WHO declared a global pandemic on March 11, 2020 (Park et al., 2020). This pandemic has caused global socioeconomic chaos, e.g., causing problems in sport, educational, religious, political, and cultural activities. So it is one the

most important problems in the world (Santos et al., 2019; Fung et al., 2019; Quwaider and Jararweh, 2016).

Besides ongoing studies about the pathogenesis of this disease (Ahmet et al., 2020; Al-Zinati et al., 2020), the accurate prediction of disease outbreaks has remained a challenge for governments and policymakers (Graham et al., 2018; Metcalf and Lessler, 2018). Therefore, predicting disease outbreaks has been found appealing to researchers (Hassani et al., 2019). In this regard, spatial aspects have been demonstrated to have a crucial role in predicting disease outbreaks (Melin et al., 2020). Geographic information system (GIS) is useful tool in the public health domain, particularly for the infectious disease surveillance and modelling strategies (Saran et al., 2020). GIS offers spatial modeling options that accounts for the influence of various factors (Abousaeidi et al., 2016). Therefore, GIS and spatial big data technologies play a pivotal role in identifying the spatial transmission in a disease outbreak (Zhou et al., 2020). Regarding the COVID-19 outbreak, Mollalu et al. have investigated the 36 environmental, socioeconomic, topographic, and demographic variables, explaining the spatial variability of disease incidence using statistical methods in ArcGIS 10.5. The results showed that the multiscale geographically weighted regression (MGWR) model could explain the highest variation and the lowest AICc compared to others (Mollalo et al., 2020). Suleman et al. have shown that GIS techniques can effectively analyze vulnerable geographical locations in Pakistan (Sarwar et al., 2020). Moreover, Melin et al. have described self-organizing maps for the spatial evolution of coronavirus pandemic, which can categorize the countries behaving similarly in dealing with the coronavirus spread; thus, these countries can benefit from similar strategies (Melin et al., 2020). Consistent with these, Zhou et al. have shown that GIS and big data technologies can play substantial roles in controlling the COVID-19 outbreak (Zhou et al., 2020).

Iran is one of the first countries that was affected by this pandemic. In Iran, the first COVID-19 case was confirmed on February 19, 2020 (WHO, 2019). Based on the Ministry of Health and Medical Education in Iran, 350279 people were infected, and 20125 people died from this disease from February 19 to August 19 (Ministry of Health and Medical Education of Iran, 2019). With the ever-increasing figures of COVID-19 cases in Iran, the government has imposed restrictions all across the country (Sarwar et al., 2020).

Herein, we aimed to investigate the relationship between demographic, economic, environmental, health infrastructure, social, and topographic variables with COVID-19 incidence and its mortality in Iran. Moreover, we intended to apply them to predict the future of COVID-19 in Iran. To the best of the authors' knowledge, it is the first study evaluating these variables to predict the COVID-19 in Iran.

2. Materials And Methods

2.1. Data collection

In this study, we surveyed the impact of 36 different variables in COVID-19 incidence and mortality through predicting its outbreak in Iran. We got the number of COVID-19 cases and the number of dead people due to it from the Ministry of Health and Medical Education of Iran.

As seen in Fig. 1, the infection and mortality rates and the mortality to infection ratio vary in different provinces. According to Fig. 2 (a), Tehran, Fars, Khorasan Razavi, Azarbayjan Shargi, Ehfahan, and Khuzestan provinces suffered from the highest infection rate in the whole country. Tehran, Khuzestan, and Azarbayjan Shargi provinces also had the highest mortality rate (Fig. 2b). In Zanjan, Hamedan, Ardebil, Golestan, and Kerman provinces, the highest mortality to infection ratio was found (Fig. 2c).

We categorized the independent variables into 6 groups. As shown in Table 1 we collected these data from different sources at the provincial level and joined them to the administrative boundary shapefile using ArcGIS Desktop 10.5.

2.2. Methods

Considering our major goals, data analysis comprised two steps. First, we examined the impacts of different variables on both the COVID-19 incidence and its mortality. Our objective was to figure out if these independent variables were effective, and if yes, what was the size of the impact on dependent variables. To achieve this, the OLS model and Adjusted R-Squared besides scatter matrix plot in ArcGIS 10.5 were employed. In addition to performing spatial analysis, the OLS model is a regression method that investigates the relationships between a set of explanatory or independent variables and dependent variables (Wheeler and Calder, 2007). So, we used this model to exclude ineffective variables and those correlated poorly with the dependent variables. In the second step, we aimed to predict the incidence of the COVID-19 infection and its associated mortality in Iran by analyzing the relationship between independent and dependent variables. So, first, we resolved the multiple collinearities between independent variables. We used the variance inflation factor (VIF) for this purpose and omitted the variables which had VIF larger than 7.5. In the next step, the ordinary least squares (OLS) and the geographical weighted regression (GWR) models were performed for non-collinear variables using ArcGIS10.5 software for spatial analysis and predicting the Covid-19 incidence and mortality.

In the ordinary spatial regression models such as the OLS, global parameters are produced to assess spatial relationships (Yuan et al., 2020). The regression parameter estimation value generated by the global regression models is the average of the whole region and has no local significance, let alone reflect the true spatial characteristics of regression parameters. Therefore, it is impossible to adequately explain an individual situation, and consequently the spatial heterogeneity, by using global overall parameters (Fotheringham et al., 2002); therefore, the findings of classical estimation models are liable to bear certain degrees of bias (Zhang et al., 2020). The details of the OLS may be found in Anselin and Arribas-Bel, 2013; Mollalo et al., 2020 and Yue et al., 2018.

Table 1
Independent variables in 6 groups

Category	Variable Name	Description	Source
Demographic	(1)Having a large population	Contribution of the variable in the country	Province statistical yearbook of 1397
	(2)Having a large number of men		
	(3)Having a large urban population		
	(4)Having a large number of people aged 65 and over		
	(5)High natural mortality		
Economic	(1)Percent of family's annual income	Contribution of the variable in the country	Province statistical yearbook of 1397
	(2)Percent of economically active people		
	(3)Unemployment rate	Percent of jobless people	
Environmental	(1) Having a high percentage of roads	Contribution of the roads in the country	Province statistical yearbook of 1397
	(2)Increasing rain Average	30-year average of rainfall	
	(3)Increasing temperature Average	30-year average of temperature	
Health infrastructure	(1)Lack of doctors	Contribution of the variable in the country	Province statistical yearbook of 1397
	(2) Lack of general practitioner		
	(3) Lack of specialist practitioner		
	(4) Lack of subspecialist physician		
	(5) Lack of paramedics		
	(6) Lack of nurse		
	(7) Lack of doctor assistant		
	(8) Lack of hospital		
	(9) Lack of hospital's bed		
	(10) Lack of ICU's bed		

Category	Variable Name	Description	Source
	(11) Lack of medical laboratory		
	(12) Lack of emergency center		
	(13) Lack of health care center		
	(14) Lack of health house		
Social	(1)Low number of insured people	Contribution of the variable in the country	Province statistical yearbook of 1397
	(2)High number of doctor consultation		
	(3)High incoming immigrants rate		
	(4)High number of annual hospitalization		
	(5) High number of incoming passengers from the other province		
	(6) High number of intra-provincial passengers		
	(7) High number of supported people by charity organizations		
	(8)High literacy rate	literacy people among over 6-year people	
	(9)High age average	Average of people's age	
Topographic	(1)Increasing altitude average	Average of height in the whole province	DEM of Shuttle Radar Topography Mission (SRTM)
	(2) Increasing the slope average	Average of slope in the whole province	

According to Tobler's first law of geography, each object is related to another object, but close objects are more related than distant objects (Worboys and Duckham, 2004). Regarding the prevalence of COVID-19, it is clear that the variables are spatially correlated and according to this, we should consider the spatial correlation between independent variables and dependent variables. Fotheringham proposes geographical weighted regression (GWR) which consider spatial heterogeneity, geographic coordinates and core function to carry out local regression estimation on adjacent subsamples of each group (Wu, 2020). The geographically weighted regression (GWR) model expands the classical regression framework that effectively addresses issues of spatial heterogeneity by enabling the variable coefficients to change

with the spatial locations (Sun and Xu, 2016). Regarding to this, one of the important aspects of GWR is focusing on the geographical location of the observations, while coefficients are locally estimated as they are allowed to vary spatially (Lykostratis and Giannopoulou, 2020). The GWR model takes the samples within a defined neighborhood into the calculation by giving more weights to nearby samples than those further away (Wheeler and Calder, 2007; Zhang et al., 2011). So the GWR model can describe the relationship between the dependent and explanatory variables and reflect spatial heterogeneity (Yu et al., 2020). This model is widely used in economics, geography, forestry, meteorology, etc. (Liu et al., 2019). The details of the GWR may be found in Fotheringham et al., 2002; Fotheringham et al., 1998 and Fotheringham and oshan, 2016.

For the OLS and GWR model accuracy evaluation, Adjusted R2 and Akaike information criterion (AIC) were used. Because the AIC approach takes differences in the freedom degree of different models into account, compared with the other methods, the AIC method can solve the problem more rapidly and conveniently (Liu et al., 2019). For more details see Yu et al., 2020; Zhang et al., 2020 and Fotheringham et al., 2002.

3. Results And Discussion

After compiling 36 independent variables, we surveyed their effectiveness on the two dependent variables using Adjusted R-Squared. Adj.R2 suggests that how an independent variable effects a dependent variable. Low Adj. R2 implies that the independent variable can't explain the dependent variable significantly, and the opposite is true for high Adj.R2. The results of this survey are demonstrated in Table 2. Accordingly, the impacts of independent variables on the COVID-19 infection and its mortality are very different. We considered the variables which their Adj.R2 were higher than 0/7 as effective ones. Thus, we found out that 20 variables out of the 36 independent variables could explain COVID-19 infection, and 16 variables could explain the mortality of the COVID-19.

Table 2 shows that the variables including having a large number of men, having a large population, lack of specialist doctors, lack of hospital, having a large urban population, having a large number of people aged 65 or older individuals, and high natural mortality rate had the most prominent impact on the COVID-19 infection increasing rate, respectively. On the other hand, increasing temperature average, increasing unemployment rate, increasing slope average, increasing number of economically active people, increasing altitude average, and increasing rainfall average had the least impact on the COIVID-19 infection increasing rate, respectively. However, the examination of the impact of independent variables on the mortality rate of the COVID-19 revealed some conflicting results; so that lack of ICU beds, low number of insured people, lack of subspecialist physicians, and lack of hospital beds had the most prominent impact on increasing of COVID-19 mortality; and lack of health houses, increasing intra-provincial travels, increasing altitude average, increasing slope average, increasing rainfall average and having high percent of roads had the least impact on the COIVID-19 infection increasing rate, respectively. Figure 3 shows effective and ineffective independent variables on the infection rate (a) and the mortality of COVID-19 (b).

In addition to examining the Adj.R2, we drew the scatter matrix plot to explore the relationship between independent and dependent variables. The results are similar to those obtained by using the Adj.R2 use. Figure 4 illustrates some of the matrix plots of effective and ineffective variables'.

Table 2
Impacts of independent variables on the COVID-19 infection and its mortality

Category	Variable Name	Infection		Death	
		R^2	Adj. R^2	R^2	Adj. R^2
Demographic	(1)Having a large population	**0.921601**	**0.918898**	**0.837471**	**0.831867**
	(2)Having a large number of men	**0.921629**	**0.918926**	**0.836103**	**0.830451**
	(3)Having a large urban population	**0.903996**	**0.900686**	**0.888763**	**0.884927**
	(4)Having a large number of people aged 65 and over	**0.901503**	**0.898107**	**0.849336**	**0.844140**
	(5)High natural mortality	**0.900438**	**0.897005**	**0.805765**	**0.799067**
Economic	(1)Percent of family's annual income	0.265130	0.239790	0.296421	0.272160
	(2)Percent of economic active people	0.004084	-0.030258	0.041465	0.008412
	(3)Unemployment rate	0.005708	-0.028578	0.010323	-0.023804
Environmental	(1)Having high percent of roads	0.081739	0.050075	0.000115	-0.034364
	(2)Increasing rain Average	0.000000	-0.034483	0.000152	-0.034326
	(3)Increasing temperature Average	0.005978	-0.028298	0.014633	-0.019346
Health infrastructure	(1)Lack of doctors	**0.858574**	**.853698**	0.689371	0.678659
	(2) Lack of general practitioner	0.426534	0.406759	0.190749	0.162843
	(3) Lack of specialist practitioner	**0.909654**	**0.906538**	**0.816437**	**0.810107**
	(4) Lack of subspecialist physician	**0.857349**	**0.852430**	**0.907467**	**0.904276**
	(5) Lack of paramedics	**0.705673**	**0.695524**	0.417355	0.397264
	(6) Lack of nurse	**0.776071**	**0.768349**	0.487775	0.470112
	(7) Lack of doctor assistant	0.295785	0.271502	0.155724	0.126611
	(8) Lack of hospital	**0.904625**	**0.901336**	**0.801735**	**0.794898**
	(9) Lack of hospital's bed	**0.892623**	**0.888921**	**0.891791**	**0.888060**

Bold cells are significantly effective on infection and death of COVID-19

	(10) Lack of ICU's bed	**0.860957**	**0.856163**	**0.924012**	**0.921392**
	(11) Lack of medical laboratory	**0.868171**	**0.863625**	**0.805765**	**0.799067**
	(12) Lack of emergency center	0.235739	0.209358	0.041465	0.008412
	(13) Lack of health care center	**0.791024**	**0.783817**	0.516292	0.499612
	(14) Lack of health house	0.092374	0.061077	0.006111	-0.028161
Social	(1)Low number of insured people	**0.888764**	**0.884928**	**0.913515**	**0.910533**
	(2)High number of doctor consultation	**0.877923**	**0.873714**	**0.856553**	**0.851606**
	(3)High incoming immigrants rate	**0.859000**	**0.854138**	**0.852076**	**0.846976**
	(4)High number of annual hospitalization	**0.835445**	**0.829771**	**0.752567**	**0.744035**
	(5) High number of incoming passengers from the other province	**0.824903**	**0.818865**	**0.825182**	**0.819152**
	(6) High number of intra-provincial passengers	0.132030	0.102100	0.002833	-0.031552
	(7) High number of supported people by charity organizations	0.238306	0.212041	0.118999	0.088620
	(8)High literacy rate	0.123405	0.093178	0.115002	0.084485
	(9)High age average	0.049790	0.017024	0.056263	0.023720
Topographic	(1)Increasing altitude average	0.003649	-0.030708	0.00083	-0.033624
	(2) Increasing slope average	0.004623	-0.029700	0.000728	-0.033730
Bold cells are significantly effective on infection and death of COVID-19					

After investigating the effectiveness of 36 independent variables and detecting the effective ones, we eliminated non-significant variables. In the next step, we aimed to predict the COIVD-19 incidence and mortality in Iran. For this purpose, we needed to solve multiple collinearity problems between independent variables, so we used variance inflation factor (VIF) to examine multiple collinearities between 20 variables for infection and 16 variables for mortality due to COVID-19. We eliminated variables with high VIF one by one until the VIF dropped to 7/5. Finally, only two variables for infection and two variables for mortality of COVID-19 were chosen to be included in the spatial analysis and prediction. Table 3 demonstrates the multiple collinearity analysis results. Accordingly, the value of VIF index is lower than 7/5. The infection variables are high incoming immigrants rate and lack of nurses and for mortality of COVID-19 were high number of doctor consultation and high incoming immigrants rate. Then we found

some other important statistical parameters in spatial analysis and prediction by the OLS and GWR models that we that merit mentioning. The coefficient was one of the important parameters representing the strength and type of the relationship between independent and dependent variables. The higher coefficient value, the better the model and the actual fitting effect; and when the coefficient was negative, the relationship is negative and when it is positive, the relationship is positive (Yu et al., 2020); So according to Table 3, the coefficient is a strong significant of both infection rate and the mortality OLS analysis. The second important parameter was the probability or p-value, which must be lower than 0.01 ($p < 0.01$) to be statistically significant (Mollalo et al., 2020); so, regarding Table 3, all of the variables' were statistically significant ($p < 0.01$). However, if the koenker test is statistically significant, we should use the robust probabilities to assess explanatory variable statistical significance; so in our OLS result the koenker test is statistically significant and the robust pr is also statistically significant ($p < 0.01$). Eventually, according to the results of the OLS, which are shown in Table 3, the statistical parameters were significant.

Table 3
VIF and important statistical parameters

	Variable	Coefficient	t-Statistic	Probability	Robust_Pr	VIF
Infection	Incoming immigrants	2618.149701	6.292509	0.000001*	0.000000*	2.939438
	Nurse	2172.058404	3.816076	0.000685*	0.001207*	2.939438
Death	Incoming immigrants	104.535586	2.991613	0.005737*	0.002231*	6.291144
	Annual doctor's visit	114.499893	3.178478	0.003596*	0.002231*	6.291144
*An asterisk next to a number indicates a statistically significant						

After examining the statistical accuracy of the selected variables, we performed OLS and GWR models to predict the COVID-19 infection and mortality rates; then we used adjusted-R2 and AICc to compare the performance of both models. The higher adjusted R2 value is, the better the model and the actual fitting effect and the smaller the ACIc value is, the better is the model fitting degree. For more details about adjusted R2 and AIC see Zhang et al., 2020 and Yu et al., 2020. Regarding Table 4, Adj.R2 for infection and death is around 0.9, but for GWR is slightly better than the OLS; also, AICc's values, for death are lesser than infection; for death analysis, it's fitter than infection analysis in both models. But the results show that AICc's values are the same in the OLS and GWR; just for death it is a bit better than in GWR and the opposite for infection. If the difference between AICs is less than around 3, the performance of models is the same (Fotheringham et al., 2002).

Table 4
Comparing OLS and GWR results

Criterion	OLS		GWR	
	Infection	death	Infection	death
Adj. R^2	0.900617	0.883533	0.902149	0.894194
AIC_C	618.817728	441.645874	620.2046	441.28969

After examining the models' propriety for predicting COVID-19, we performed GWR and OLS models on the spatial data. As seen in Fig. 5, the dark red areas depict areas where the actual values are higher than where the model predicted. On the contrary, the light blue to dark blue values indicate where the actual values are lower than the model predicted. Figure 5(a) and (b) are respectively the OLS and GWR models' results in COVID-19 infection that have mostly the same results in different provinces excluding Tehran and Ilam. Regarding Fig. 5 and Table 5, COVID-19 infection will increase in Kerman, Kermanshah, Alborz, Hamedan, Khorasan Jonubi, Khorasan Razavi, Esfahan, and Semnan; On the contrary, it will decrease in Lorestan, Khuzestan, Azarbayjan Shargi, Fars, Golestan, Hormozgan, Kordestan, Mazandaran and Sistan va Baluchestan. However, the degree of fluctuations varies in different provinces. There was not any change in the other provinces, and they will continue the previous process. These results are common to both models. But according to the GWR result, infection in Tehran and Ilam will decrease, while in the OLS result, their infection process won't change. As seen in Fig. 6 and Table 5, Tehran, Esfahan, Khorasan Razavi, and Fars will have the largest infection contribution, and Khorasan Shomali, Ilam, Golestan, and ChaharmahaloBakhtiari will have the lowest infection contribution.

Mortality due to COVID-19 prediction results is shown in Fig. 7 and Table 5. According to these, mortality will increase in Fars, Khorasan Razavi, Alborz, and Esfahan, and it will decrease in Tehran, Zanjan, Lorestan, Khuzestan, Hormozgan, Golestan, Gilan, Bushehr, and Ardebil. But the intensity of the increase or decrease varies in the different provinces. In some of the provinces, the models' results are different. But the main difference is in Azarbayjan Gharbi, the OLS predicts it will decrease and GWR predicts it will increase. According to Fig. 8 and Table 5, Tehran, Esfahan, Fars, Khuzestan, Khorasan Shomali, Alborz, and Azarbayjan Shargi will have the largest mortality contribution and Ilam, ChaharmahaloBakhtiari, Kohgiluye va Boyerahmad, and Semnan, will have the lowest COVID-19 mortality contribution.

Table 5
Prediction of COVID-19 infection and mortality due to it at the provincial level

Province Name	Infection				Death			
	Type Of Change		Prediction		Type Of Change		Prediction	
	OLS	GWR	OLS	GWR	OLS	GWR	OLS	GWR
Alborz	+ 1	+ 1	19615	19015	+ 1	+ 1	1084	1126
Ardebil	0	0	6564	6593	-1	-1	298	321
Azarbayjan Gharbi	0	0	14640	15166	-1	+ 1	687	734
Azarbayjan Shargi	-2	-2	18905	19761	0	0	1314	1375
Bushehr	0	0	6555	6442	-1	-1	245	223
ChaharmahaloBakhtiari	0	0	5156	4945	0	0	194	178
Esfahan	+ 1	+ 2	33305	33418	+ 1	+ 1	1295	1282
Fars	-1	-1	32428	32766	+ 2	+ 2	1304	1260
Gazvin	0	0	8027	7720	0	0	254	263
Gilan	0	0	16321	16460	-1	-1	554	584
Golestan	-1	-1	4384	4367	-1	-1	422	401
Hamedan	+ 1	+ 1	8560	8601	0	0	413	441
Hormozgan	-1	-1	9548	9666	-1	-1	464	432
Ilam	0	-1	2470	2220	0	0	104	117
Kerman	+ 2	+ 2	13755	12824	+ 1	0	712	643
Kermanshah	+ 2	+ 2	11923	12504	0	0	538	475
Khorasan Jonubi	+ 1	+ 1	5516	5124	+ 1	0	306	235
Khorasan Razavi	+ 1	+ 1	37115	33821	+ 2	+ 2	1506	1403
Khorasan Shomali	0	0	4689	4022	0	0	318	285
Khuzestan	-2	-2	18588	19389	-1	-1	1015	1094
Kohgiluye va Boyerahmad	0	0	5156	5094	0	0	141	104
Kordestan	-1	-1	10235	10080	0	+ 1	538	578
Lorestan	-2	-2	8442	8890	-1	-1	282	301
Markazi	0	0	7161	6944	0	+ 1	461	489

Province Name	Infection				Death			
	Type Of Change		Prediction		Type Of Change		Prediction	
	OLS	GWR	OLS	GWR	OLS	GWR	OLS	GWR
Mazandaran	-1	-1	16618	16648	0	0	710	747
Qom	0	0	6705	6405	0	0	321	322
Semnan	+ 1	+ 1	6463	5921	+ 1	0	237	170
Sistan va Baluchestan	-1	-1	9158	8717	0	-1	461	404
Tehran	0	-2	74614	74418	-2	-2	4136	4344
Yaz	0	0	7526	7032	0	-1	398	335
Zanjan	0	0	5338	5174	-1	-1	252	272

Meaning of *type of change's* sign: +2: will increase a lot, + 1: will increase, 0: no change, -1: will decrease, and − 2: will decrease a lot.

4. Conclusion

According to this article's main purposes, we investigated 36 independent variables' impact on the COVID-19 incidence and the mortality due to it as two dependent variables in Iran. We categorized these independent variables into 6 different groups to determine what kind of variables has the most effects on the dependent variables. Then we analyzed the COVID-19 spatially to predict its incidence and mortality in the future in Iran. For this purpose, we used the OLS and GWR models.

The results indicated that 20 different variables had a high correlation with the COVID-19 incidence in Iran. Five of these variables were demographic, 10 were health infrastructure, and 5 were social variables. So economical, environmental, and topographical variables such as increasing temperature average, increasing altitude average, annual household's income average, unemployment rate, etc., don't affect the dependent variable, or their effects are very low. Examining of the impact of independent variables on the COVID-19's mortality showed similar results as its incidence results except for an interesting difference; the most influential variables on the COVID-19 incidence are demographic indicators, but the most influential variables on the mortality due to COVID-19 are health infrastructure such lack of ICU beds, lack of subspecialist physicians, lack of hospital beds, etc. Eventually we can recognize that the health infrastructure plays a very important role in COVID-19 incidence and its mortality.

In the other hand, we analyzed the COVID-19 spatially to predict its infection and mortality rates in the future across the different provinces. We used the OLS and GWR models for prediction. Generally, the results indicated that the COVID-19 infection would have a considerable increase in Kerman, Esfahan, and Kermanshah, and there would be a significant reduction in Khuzestan, Lorestan, Azarbayjan Shargi, and Tehran. In addition, the mortality due to COVID-19 is expected to increase considerably in Fars and

Khorasan Razavi and, at the same time, would decrease significantly in Tehran, Ardebil, Zanjan, Gilan, Golestan, Lorestan, Khuzestan, Bushehr, and Hormozgan. Regarding the results, Tehran, Esfahan, Fars, and Khorasan Razavi would have the highest, and Ilam, Khorasan Shomali, and Golestan would have the lowest COVID-19 infection rates. Also, Tehran, Esfahan, Fars, Khorasan Razavi, Khuzestan, and Azarbayjan Shargi would have the highest, and Ilam, ChaharmahaloBakhtiari, Kohgiluye va Boyerahmad, and Semnan will are expected to have the lowest mortality due to COVID-19.

Declarations

Ethical Approval and Consent to participate

Not applicable

Consent for publication

The authors are satisfied to publish the article in the Journal

Availability of supporting data

Data is available upon a request via email to the first author

Competing interests

Not applicable

Funding

Not applicable

Authors' contributions

A.G., and B.B., conceptualized the work and supervised the project. H.A., and M.A., have written the manuscript. The final version is confirmed by all of the authors.

Acknowledgements

Not applicable

Authors' information

Abolfazl Ghanbari*

Faculty of planning and Environmental Sciences, University of Tabriz, Iran, danizly1358@yahoo.com

Behzad Baradaran

Immunology Research Center, Tabriz University of Medical Sciences, Iran, behzad_im@yahoo.com

Hamed Ahmadi

Faculty of Geography, University of Tehran, Iran, hamed.ahmadi1989@ut.ac.ir

Maryam Ahmadi

Faculty of planning and Environmental Sciences, University of Tabriz, Iran, maryam.ahmadi7694@gmail.com

References

1. Abousaeidi, M, Fauzi R, Muhamad R. Geographic Information System (GIS) modeling approach to determine the fastest delivery routes. Saudi Journal of Biological Sciences. 2016; 23 (5): 555-564. https://doi.org/10.1016/j.sjbs.2015.06.004.
2. Ahmed W, Angel N, Edson J, Bibby K, Bivins A, O'Brien J, Choi P, Kitajima M, Simpson S, Li J, Tscharke B, Verhagen R, Smith W, Zaugg J, Dierens L, Hugenholtz P, Thomas K, Mueller J. First confirmed detection of SARSCoV-2 in untreated wastewater in Australia: a proof of concept for the wastewater surveillance of COVID-19 in the community. Science of the Total Environment. 2020; 728(138764): 1-8. https://doi.org/10.1016/j.scitotenv.2020.138764.
3. Al-Zinati M, Almasri T, Alsmirat M, Jararweh Y. Enabling multiple health security threats detection using mobile edge computing. Simulation Model Practice and Theory. 2020; 101(101957):1-15. https://doi.org/10.1016/j.simpat.2019.101957.
4. Anselin L, Arribas-Bel D. Spatial fixed effects and spatial dependence in a single cross-section. Regional Science Association International. 2013; 92(1):3–17. https://doi.org/10.1111/j.1435-5957.2012.00480.x.
5. Fotheringham A, Brunsdon C, Charlton M. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. Environment and Planning A: Economy and Space. 1998; 30(11):1905–1927. https://doi.org/10.1068%2Fa301905.
6. Fotheringham A, Oshan T. Geographically weighted regression and multicollinearity: dispelling the myth. J. Geogr. Syst. 2016; 18(4): 303–329. https://doi.org/10.1007/s10109-014-0199-6.
7. Fotheringham S, Brunsdom C, Charlton M. Geographically Weighted Regression, the analysis of spatially varying relationships. University of Newcastle, UK.
8. Fung V, Overhage L, Sylvia L, Reilly-Harrington N, Kamali M, Gao K, Calabrese J. Complex polypharmacy in bipolar disorder: Side effect burden, adherence, and response predictors. Journal of affective disorders. 2019; 257(1):17–22. https://doi.org/10.1016/j.jad.2019.06.050.
9. Graham M, Suk J, Takahashi S, Metcalf C, Jimenez A, Prikazsky V, Ferrari M, Lessles J. Challenges and opportunities is disease forecasting in outbreak setting: a case study of measles in lola

prefecture, guinea. Am J Trop Med Hyg. 2018; 98(5):1489-1497. https://doi.org/10.4269/ajtmh.17-0218.

10. Hassani H, Yeganegi MR, Sirimal Risk management, signal processing and econometrics: A new tool for forecasting the risk of disease outbreaks. Theoretical Biology. 2019; 467(1): 57-62. https://doi.org/10.1016/j.jtbi.2019.01.032.
11. Liu Ch, Wu X, Wang L. Analysis on land ecological security change and effect factors using RS and GWR in the Danjiangkou Reservoir area, China. Journal of Applied Geography. 2019; 105(1)1-14. https://doi.org/10.1016/j.apgeog.2019.02.009.
12. Lykostratis K, Giannopoulou M. From Global to Local: GWR as an Exploratory tool for Spatial Phenomena. Nova Science Publishers, New York. 2020; 119-162
13. Manjo M.G, Satheesh Kumar M.K, Valsaraj K.T, Sivan C, Vijayan Soumya k. Potential link between compromised air quality and transmission of the novel corona virus (SARS-COV-2) in affected areas. Environmental Research. 2020; 190 (110001): 1-8. https://doi.org/10.1016/j.envres.2020.110001.
14. Melin P, Monica JC, Sanchez D, Castillo O. Analysis of Spatial Spread Relationships of Coronavirus (COVID-19) Pandemic in the World using Self Organizing Maps. Chaos, Solitons and Fractals, Nonlinear Science, and Nonequilibrium and Complex Phenomena. 2020; 138(109917):1-7. https://doi.org/10.1016/j.chaos.2020.109917.
15. Metcalf C, Lessler J. Opportunities and challenges in modeling emerging infectious disease. Science. 2018; 357(6347):149-152. https://doi.org/10.1126/science.aam8335.
16. Ministry of Health and Medical Education of Iran. Coronavirus Disease 2019 (COVID-19) Situation in Iran (Islamic Republic of). 2020. Retrieved from. https://behdasht.gov.ir Accessed 19 August 2020.
17. Mollalo A, Alimohammadi A, Shirzadi M, Malek M. Geographic information system-based analysis of the spatial and spatio-temporal distribution of zoonotic cutaneous leishmaniasis in Golestan Province, north-east of Iran. Zoonoses Public Health. 2015; 62(1):18–28. https://doi.org/10.1111/zph.12109.
18. Mollalo A, Khodabandehloo E. Zoonotic cutaneous leishmaniasis in northeastern Iran: a GIS-based spatio-temporal multi-criteria decision-making approach. Epidemiology & Infection. 2016; 144(10):2217–2229. https://doi.org/10.1017/s0950268816000224.
19. Mollalo A, Mao L, Rashidi P, Glass G. A GIS-based artificial neural network model for spatial distribution of tuberculosis across the continental United States. Environmental Research and Public Health. 2019; 16(1):157. https://dx.doi.org/10.3390%2Fijerph16010157.
20. Mollalo A, Vahedi B, Kirara M. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. Science of the Total Environment. 2020; 728(138884):1-8. https://doi.org/10.1016/j.scitotenv.2020.138884.
21. Park M, Thwaites RS, Openshaw PJM. COVID-19: Lessons from SARS and MERS. Eur J Immunol. 2020; 00:1-9.
22. Quwaider M, Jararweh Y. A cloud supported model for efficient community health awareness. Pervasive and Mobile Computing. 2016; 28(1):35–50. https://doi.org/10.1016/j.pmcj.2015.07.012.

23. Santos J, Matos T, Mendes R, Freitas C, Leite A, Rodrigues I. Influência do planejamento reprodutivo e da satisfação materna com a descoberta da gravidez na qualidade da assistência pré-natal no Brasil. Revista Brasileira de Saúde Materno Infantil. 2019; 19 (3): 537–543. http://dx.doi.org/10.1590/1806-93042019000300003.
24. Saran S, Singh P, Kumer V, Chauhan P. Review of Geospatial Technology for Infectious Disease Surveillance: Use Case on COVID-19. Journal of the Indian Society of Remote Sensing. 2020; 48(8): 1121-1138. https://doi.org/10.1007/s12524-020-01140-5.
25. Sarwar S, Waheed SS, Sarwar S, Khan A. COVID-19 challenges to Pakistan: Is GIS analysis useful to draw solutions. Science of the Total Environment. 2020; 730(139089):1-7. https://doi.org/10.1016/j.scitotenv.2020.139089.
26. Sun K, Xu Z. The impacts of human driving factors on grey water footprint in China using a GWR model. Geographical Research. 2016; 35(1):37-48. https://doi.org/10.11821/dlyj201601004.
27. Ward M.D, Gleditsch K.S. Spatial regression models. Sage Publications. 2018; 155.
28. Wheeler D, Calder C. An assessment of coefficient accuracy in linear regression models with spatially varying coefficients. Journal of Geographical Systems. 2007; 9(1):145-166. https://doi.org/10.1007/s10109-006-0040-y.
29. Worboys M, Duckham M. GIS computing perspective. CRC Press Publications. 2004; 197.
30. World Health Organization (WHO (a)). Coronavirus Disease 2019 (COVID-19) Situation in Iran (Islamic Republic of). https://covid19.who.int/region/emro/country/ir Accessed 20 March 2020.
31. Wu D. Spatially and Temporally Varying Relationships between Ecological Footprint and Influencing Factors in China's Provinces Using Geographically Weighted Regression (GWR). Journal of Cleaner Production. 2020; 261(121089):1-14. https://doi.org/10.1016/j.jclepro.2020.121089.
32. Yu H, Gong H, Chen B, Liu K, Gao M. Analysis of the influence of groundwater on land subsidence in Beijing based on the geographical weighted regression (GWR) model. Science of the Total Environment. 2020; 738(139405):1-60. https://doi.org/10.1016/j.scitotenv.2020.139405.
33. Yuan Y, Cave M, Xu H, Zhang Ch. Exploration of spatially varying relationships between Pd and Al in urban soils of London at the regional scale using geographically weighted regression (GWR). Journal of Hazardous Materials. 2020; 393(122377):1-11. https://doi.org/10.1016/j.jhazmat.2020.122377.
34. Yue Y, Sun J, Liu X, Ren D, Liu Q, Xiao X, Lu L. Spatial analysis of dengue fever and exploration of its environmental and socio-economic risk factors using ordinary least squares: A case study in five districts of Guangzhou Coty, China, 2014. International Journal of Infectious Diseases. 2018; 75(2018):39–48. https://doi.org/10.1016/j.ijid.2018.07.023.
35. Zhang C, Tang Y, Xu X, Kiely G. Towards spatial geochemical modelling: use of geographically weighted regression for mapping soil organic carbon contents in Ireland. Appl. Geochem. 2011; 26(7)1239-1248. https://doi.org/10.1016/j.apgeochem.2011.04.014.
36. Zhang P, Yang D, Zhang Y, Li Y, Liu Y, Cen Y, Zhang W, Geng W, Rong T, Liu Y, Shao J. Re-examining the drive forces of China's industrial wastewater pollution based on GWR model at provincial level.

Journal of Cleaner Production. 2020; 262(121309):1-28. https://doi.org/10.1016/j.jclepro.2020.121309.

37. Zhou Ch, Su F, Pei T, Zhang A, Du Y, Luo B, Gao Zh, Wang J, Yuan W, Zhu Y, Song C, Chen J, Xu J, Li F, Ma T, Jiang L, Yan F, Yi J, Hu Y, Liao Y, Xiao H. COVID-19: Challenges to GIS with Big Data. Geography and Sustainability. 2020; 1(1):77-87. https://doi.org/10.1016/j.geosus.2020.03.005.

Figures

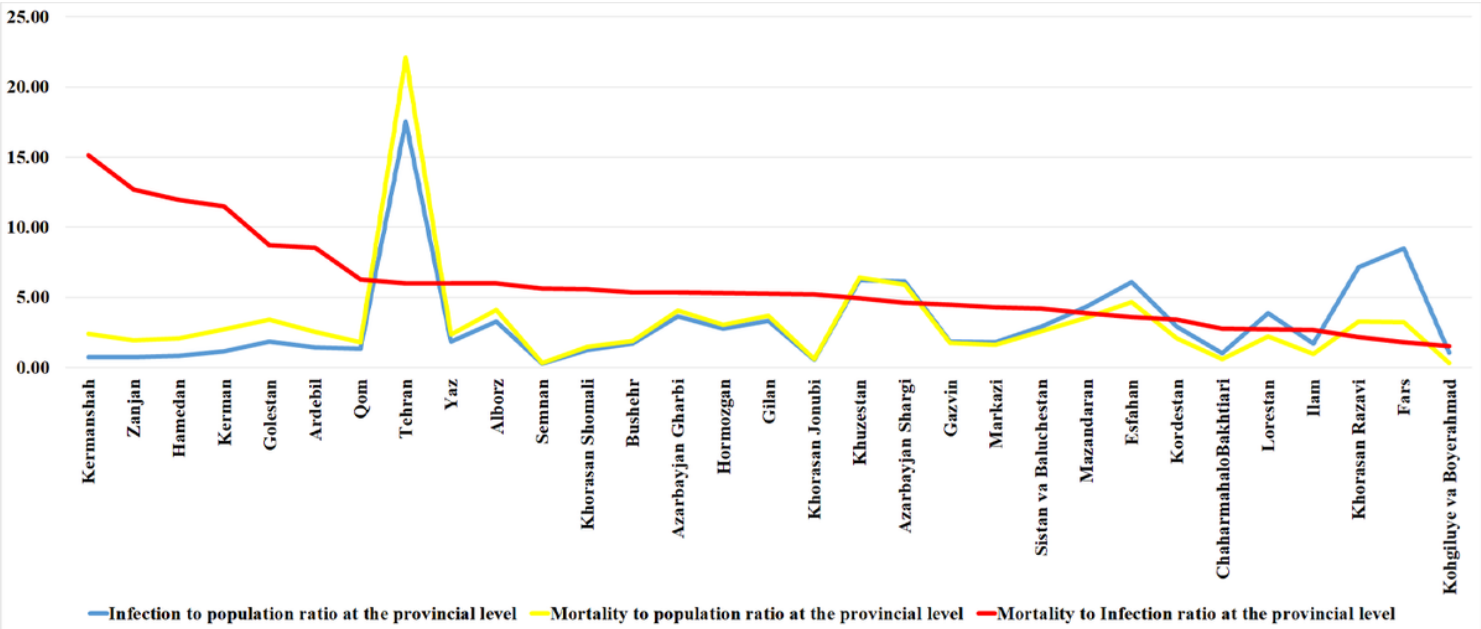

Figure 1

Infection ratio, Mortality ratio and Mortality to Infection ratio at the provincial level

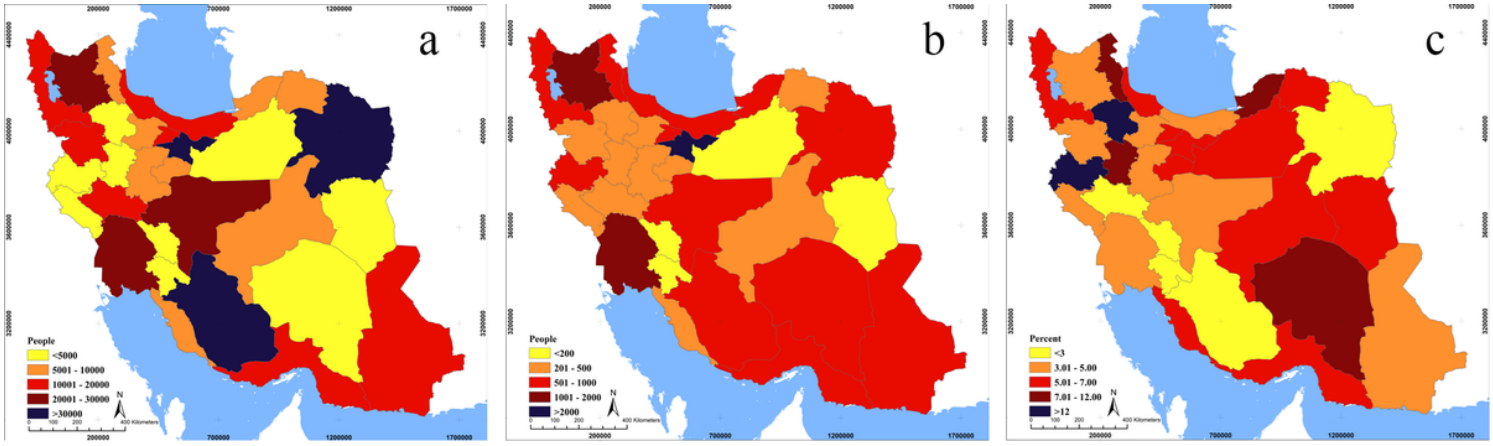

Figure 2

a) Number of infected people, b) Number of died people due to COVID-19 and c) Mortality to infection ratio Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any

country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

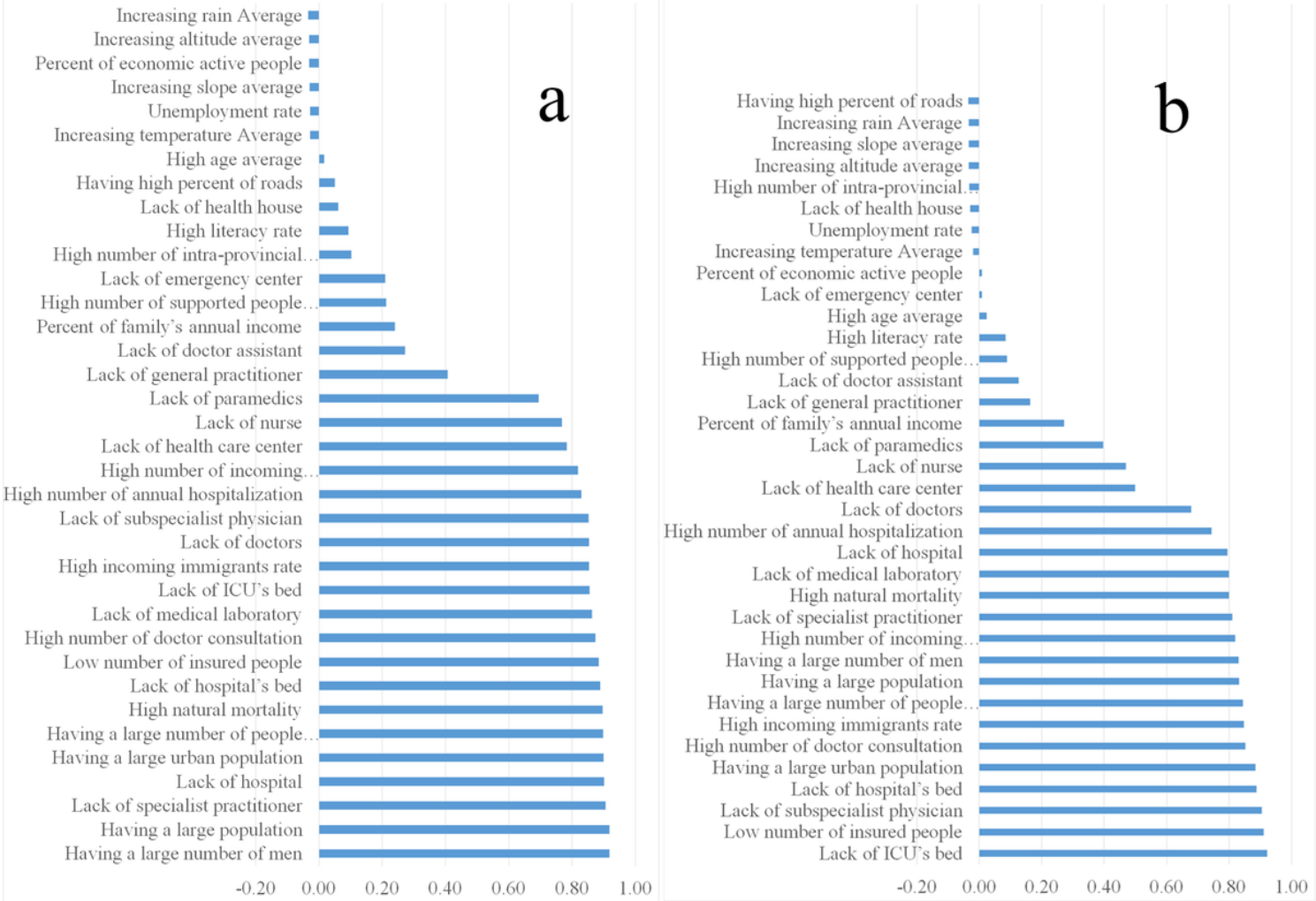

Figure 3

Effective and ineffective independent variables on a) the infection rate and b) the mortality of COVID-19

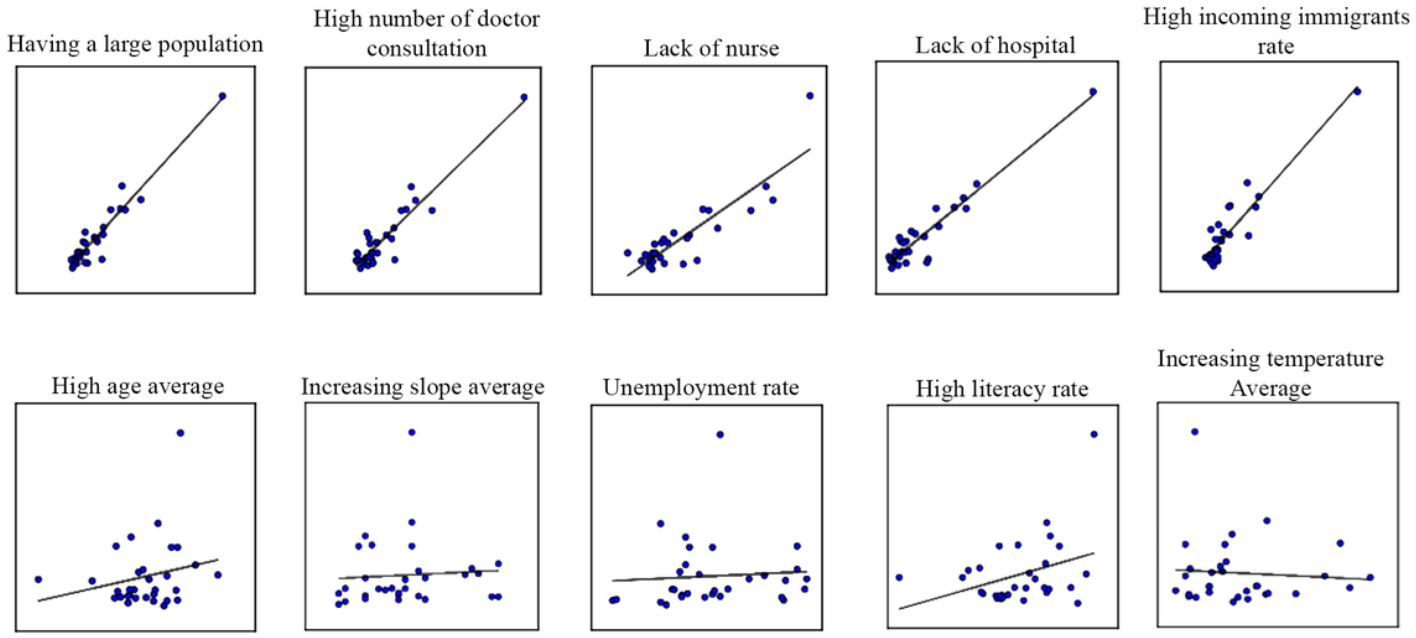

Figure 4

Scatter matrix plot for some of the effective and ineffective independent variables

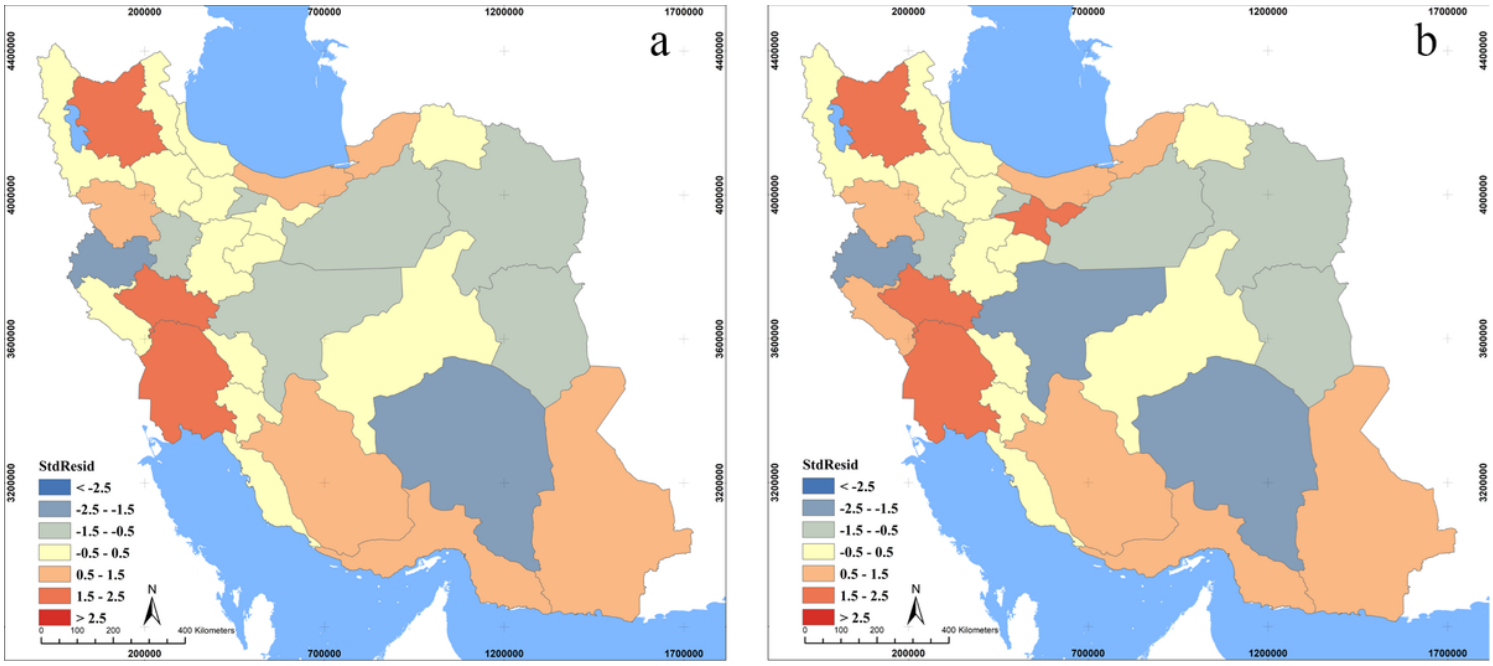

Figure 5

StdResid of a) OLS and b) GWR models for COVID-19 infection Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

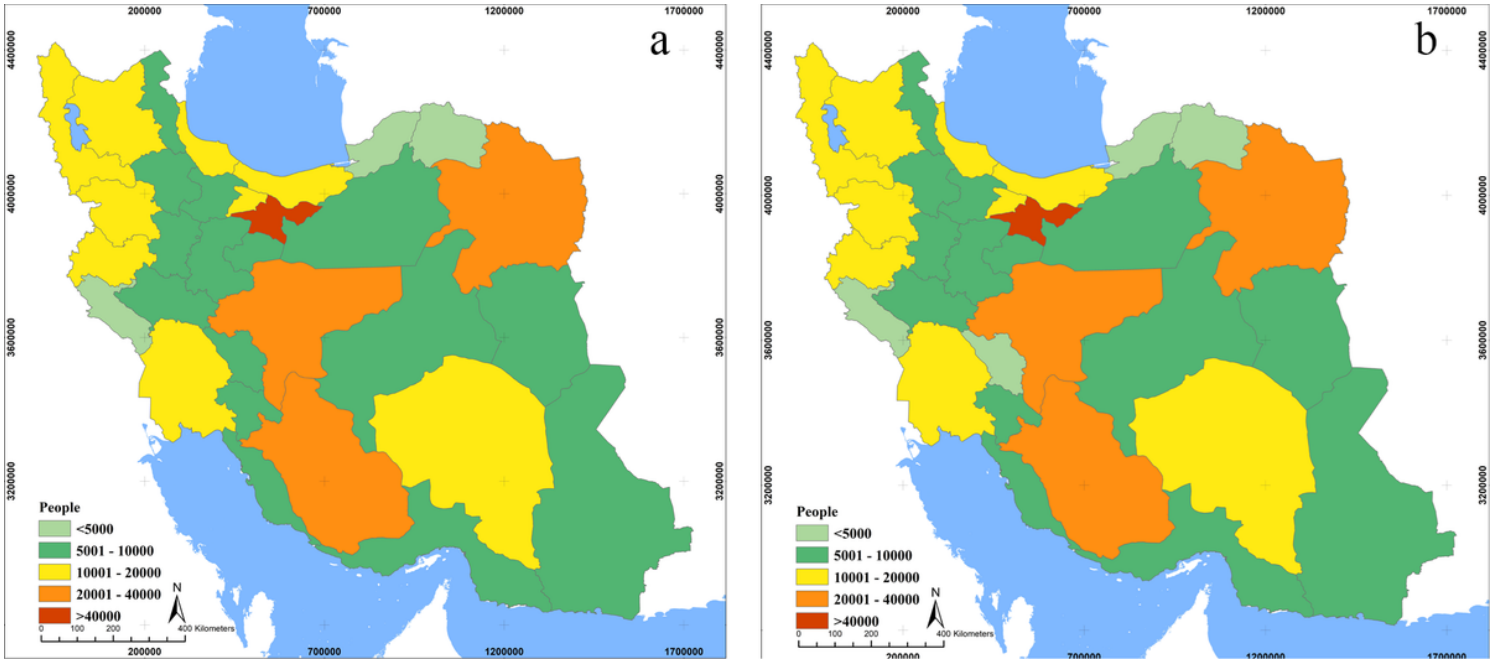

Figure 6

Prediction of COVID-19 infection a) OLS and b) GWR Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

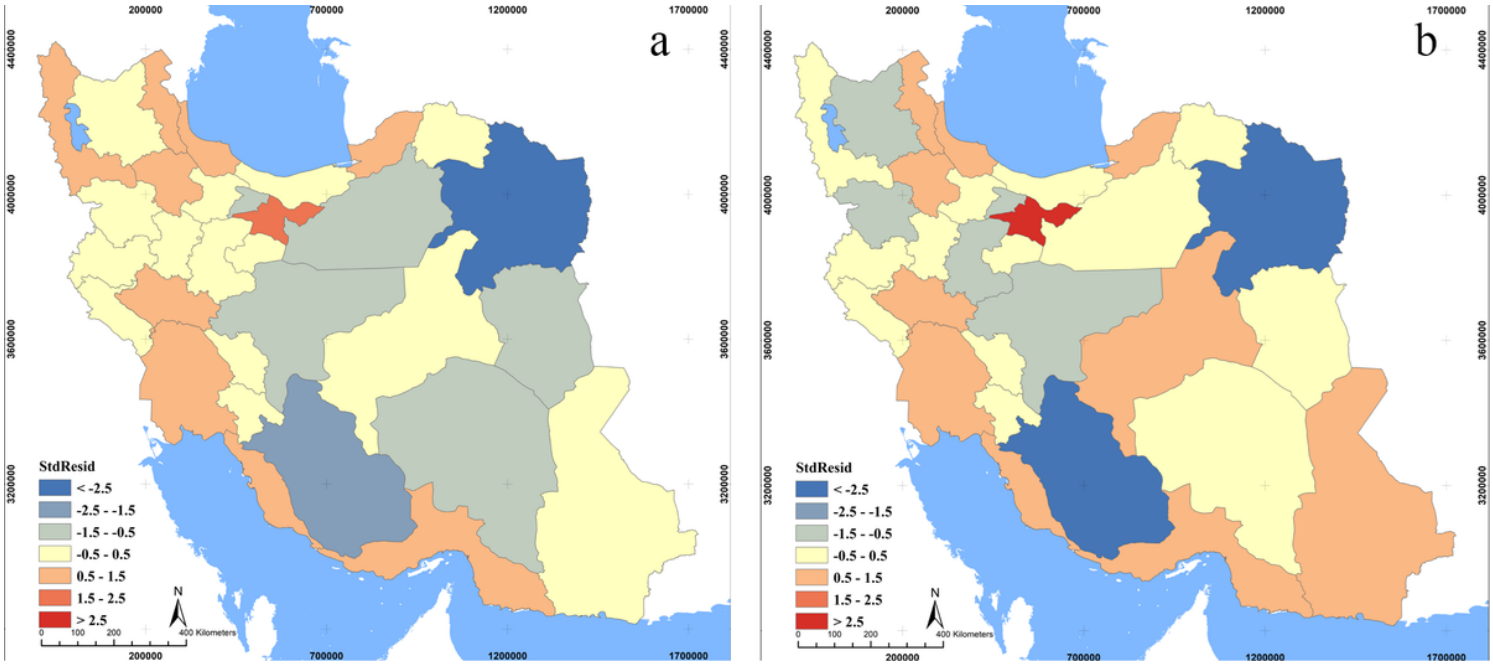

Figure 7

StdResid of a) OLS and b) GWR models for mortality due to COVID-19 Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever

on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

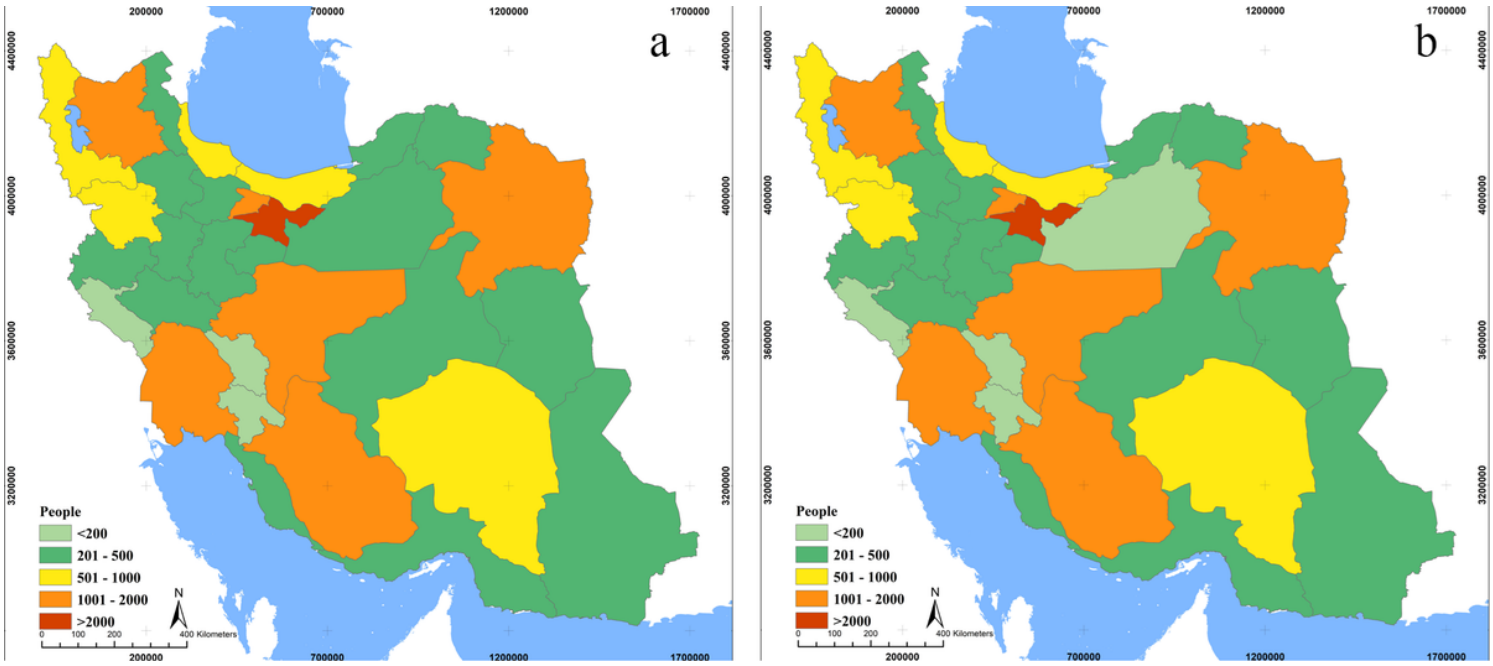

Figure 8

Prediction of mortality due to COVID-19 a) OLS and b) GWR Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.