



OPEN

## Socioeconomic and environmental determinants of foot and mouth disease incidence: an ecological, cross-sectional study across Iran using spatial modeling

Mahdi Nazari Ashani<sup>1</sup> , Ali Asghar Alesheikh<sup>1</sup> , Zeinab Neisani Samani<sup>1</sup>, Aynaz Lotfata<sup>2</sup> , Sayeh Bayat<sup>3,4</sup> , Siamak Alipour<sup>1</sup> & Benyamin Hoseini<sup>5,6</sup> ✉

Foot-and-mouth disease (FMD) is a highly contagious animal disease caused by a ribonucleic acid (RNA) virus, with significant economic costs and uneven distribution across Asia, Africa, and South America. While spatial analysis and modeling of FMD are still in their early stages, this research aimed to identify socio-environmental determinants of FMD incidence in Iran at the provincial level by studying 135 outbreaks reported between March 21, 2017, and March 21, 2018. We obtained 46 potential socio-environmental determinants and selected four variables, including percentage of population, precipitation in January, percentage of sheep, and percentage of goats, to be used in spatial regression models to estimate variation in spatial heterogeneity. In our analysis, we employed global models, namely ordinary least squares (OLS), spatial error model (SEM), and spatial lag model (SLM), as well as local models, including geographically weighted regression (GWR) and multiscale geographically weighted regression (MGWR). The MGWR model yielded the highest adjusted  $R^2$  of 90%, outperforming the other local and global models. Using local models to map the effects of environmental determinants (such as the percentage of sheep and precipitation) on the spatial variability of FMD incidence provides decision-makers with helpful information for targeted interventions. Our findings advocate for multiscale and multidisciplinary policies to reduce FMD incidence.

Foot and mouth disease (FMD) has been recognized as a significant epidemic with severe implications for the livestock industry since the sixteenth century<sup>1</sup>. FMD virus spreads through direct or indirect contact with infected animals' secretions or excretions, inhalation, ingestion, skin wounds, and mucous membranes. It causes blisters in and around the mouth, feet, and udders of animals, as well as fever, lameness, loss of condition, milk drop in dairy cattle, and sudden death in young stock due to myocarditis<sup>2</sup>. The FMD virus poses a substantial risk, often leading to widespread outbreaks because of its highly contagious nature. The 2001 FMD outbreak in England is an example, where approximately 6.7 million animals were slaughtered in the United Kingdom and the Netherlands, resulting in significant economic losses amounting to around 3.2 billion euros<sup>3,4</sup>.

While FMD has been eradicated in some parts of the world, including North America and Europe, it remains endemic in some areas in Asia, Africa, and the Middle East. Most of the economic burden of this infectious epidemic in these countries was borne by lost trade and large numbers of infected animals<sup>2</sup>. The global distribution of FMD virus serotypes is categorized into seven pools. Pools 1 and 2 encompass Southeast and Southern Asia, respectively, while Pool 3 includes Euro-Asia (including the Middle East). Pools 1, 2, and 3 collectively harbor

<sup>1</sup>Department of Geospatial Information Systems, Faculty of Geodesy and Geomatics Engineering, K. N. Toosi University of Technology, Tehran, Iran. <sup>2</sup>Department of Pathology, Microbiology, and Immunology, School of Veterinary Medicine, University of California, Davis, USA. <sup>3</sup>Department of Biomedical Engineering, University of Calgary, Calgary, AB, Canada. <sup>4</sup>Department of Geomatics Engineering, University of Calgary, Calgary, AB, Canada. <sup>5</sup>Pharmaceutical Research Center, Pharmaceutical Technology Institute, Mashhad University of Medical Sciences, Mashhad, Iran. <sup>6</sup>Department of Medical Informatics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran. ✉email: binyamin.hoseini@gmail.com

three circulating FMD virus serotypes O, A, and Asia 1. The remaining four pools, namely Pool 4, Pool 5, Pool 6, and Pool 7, represent other regions worldwide where the FMD virus is prevalent<sup>5</sup>.

Recent studies<sup>6–8</sup> have revealed that various factors may contribute to the severity of FMD. Arman et al.<sup>6</sup> identified a higher number of FMD epidemics during the driest season in Ethiopia, suggesting that seasonal changes may be a risk factor for FMD in the region. Jiang et al.<sup>7</sup> studied the climate change impacts on the FMD risk in elephants using Maximum Entropy (MaxEnt). Their study revealed that temperature seasonality, average annual temperature, average temperature in the driest quarter, and precipitation in the driest month were the most significant factors influencing FMD risk. A study conducted in Ethiopia by Udahemuka et al.<sup>8</sup> showed that vaccinating calves younger than 12 months effectively control FMD within the herd. These findings contribute to our understanding of the risk factors associated with FMD and offer valuable insights for developing effective control and prevention measures.

Spatial analysis is a valuable tool for studying the distribution of infectious diseases<sup>9–11</sup>. Ballard and Boone<sup>12</sup> used geographically weighted regression (GWR) to examine the relationship between Lyme disease and land cover in the Midwestern and Northeastern states of the United States. Some studies have used spatial analysis to model FMD<sup>13–17</sup>. Sangrat et al.<sup>13</sup> utilized a multi-criteria decision analysis method based on geographic information system (GIS) to predict FMD-prone areas in Thailand, identifying regions such as the northern, northeastern, western, and central parts of the country as FMD hotspots. Gao et al.<sup>14</sup> investigated the spatial distribution and risk zones of FMD in mainland China, identifying significant factors such as road density, isothermality, UV-B radiation seasonality, minimum temperature in the coldest month, market distribution, and railway density. Hagerman et al.<sup>15</sup> studied seasonal and geographic conditions favorable to the airborne spread of FMD in the United States, highlighting high-risk geographical areas, particularly during seasons with favorable weather conditions. These studies demonstrate the significance of spatial analysis in understanding and managing the spatial patterns and risks associated with infectious diseases.

Spatial analysis has been employed in a few studies to model FMD in Iran<sup>16,17</sup>. Ilbeigi et al.<sup>18</sup> identified Razavi Khorasan province as having a high risk of FMD prevalence due to a lack of regular on-farm hygiene control. They found that the lack of biosecurity was 11 times higher in FMD-infected farms compared to non-infected farms<sup>18</sup>. Jafarzadeh et al.<sup>16</sup> used a zero-inflated negative binomial model to estimate the probability of a "no outbreak" state and the number of outbreaks in Iran at the province level, utilizing data from 5707 FMD cases recorded between April 1995 and March 2002. Similarly, Perez et al.<sup>17</sup> employed spatial scan statistics to examine the spatial distribution of FMD in Iran using 4477 FMD cases reported from June 1996 to September 2003. Their analysis identified significant clusters of FMD that coincided with road networks, neighboring countries, and densely populated areas<sup>17</sup>.

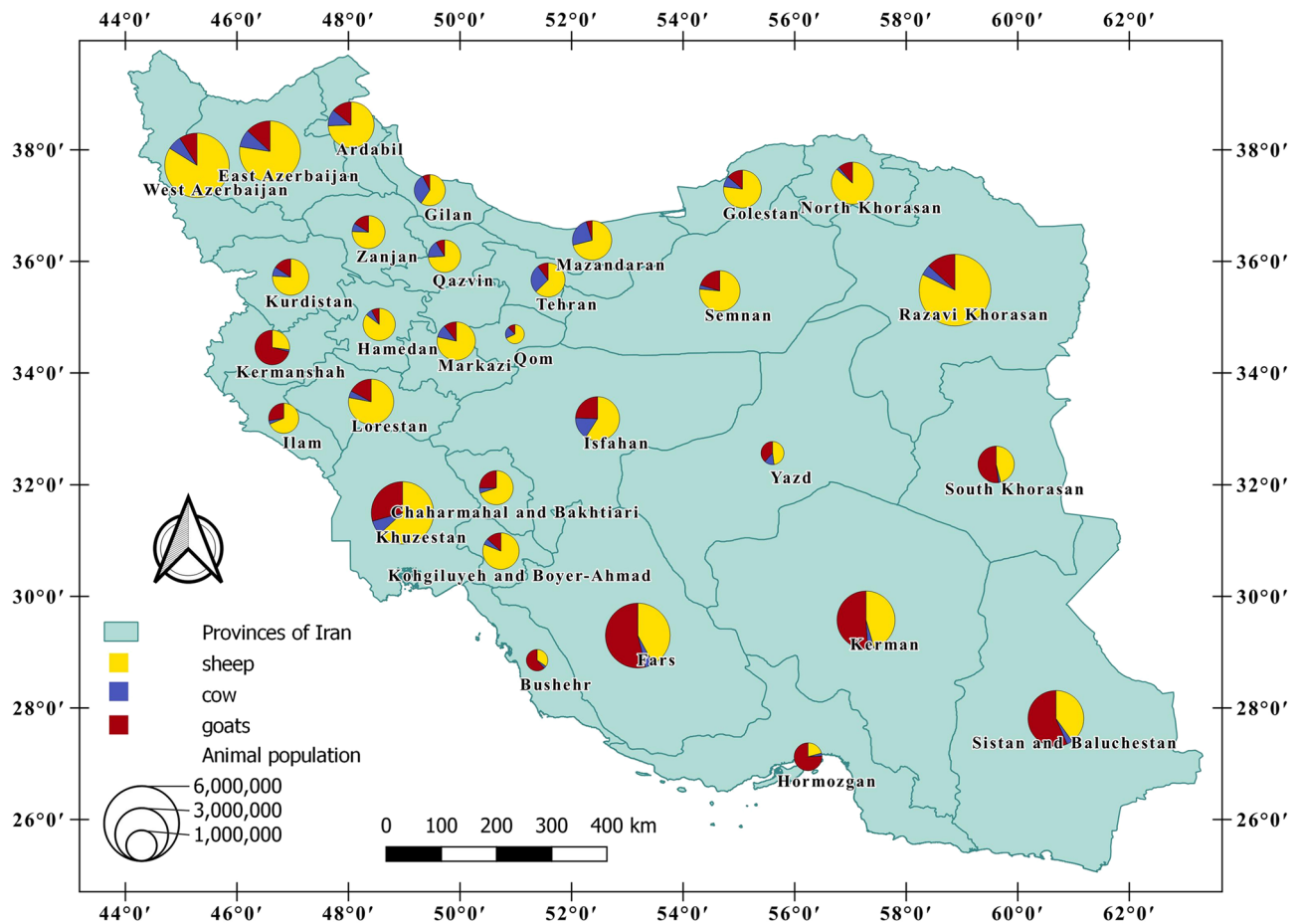
This study investigates the determinants and their significance in influencing provincial-level FMD incidence across Iran. The study's objectives encompass several aspects, including conducting exploratory analysis to identify the determinants of FMD incidence, examining the geographical variations of these determinants using local spatial modeling techniques, and analyzing FMD spatial variations through both aspatial and spatial regression models. Notably, this research represents the first study to compare local modeling with global modeling approaches in assessing the distribution of FMD across Iran. By addressing these objectives, this study provides valuable insights into the factors driving FMD incidence at the provincial level. It contributes to a better understanding of the spatial patterns of FMD in Iran.

## Methods

**Study area.** Iran, located in the Middle East (32.4279° N, 53.6880° E), is the focal point of this study as a representative case study within Asia. The country is divided into 31 provinces, each characterized by varying precipitation levels, temperature ranges, and diverse climates, including hot, dry, moderate, and humid conditions. Iran boasts a diverse topography, ranging from regions below sea level to mountainous areas reaching heights of approximately 5600 m above sea level<sup>19</sup>. The average province size is around 20,871.9197 square miles, with an average population of 2,573,796. Provinces were selected as the analytical units for this study, considering data availability and relevance. Figure 1 visually depicts the animal population distribution across Iran's provinces, providing valuable insights into the spatial patterns of animal population density.

**FMD incidence as the outcome variable.** Animal FMD cases for 30 census provinces in Iran (see Fig. 2) were obtained from the Iran Veterinary Organization (IVO). The province of Alborz was excluded from the analysis due to insufficient data availability. For this study, the outcome variable was the incidence of FMD calculated as the number of cases per 1,000,000 animals in Iran's provinces between March 21, 2017, and March 21, 2018.

**Covariates.** In this study, we evaluated 46 potential socio-environmental determinants (see Table 1). Among these determinants, we selected four variables, namely the percentage of population, precipitation in January, percentage of sheep, and percentage of goats, to be used as covariates in the spatial regression models. The data for these variables were sourced from multiple reliable sources to ensure comprehensive coverage and maintain the accuracy of our analyses. In the first step, we obtained data on the percentage of the population, educated individuals, and migrants from the Population and Housing Census of Iran<sup>20</sup>. The unemployment rate per province was acquired from The Summary of the Results of the 2016 Workforce Statistics Plan of Iran<sup>21</sup>. In the second step, we collected data on the percentage of sheep, goats, and cows per province from the National Agricultural Census of Iran<sup>22</sup>. Third, we gathered data on the percentage of the Gross Domestic Product (GDP) per province from the Iran Centre of Statistics Ministry of Cooperatives, Labor, and Social Welfare for the period between March 21, 2017, and March 21, 2018<sup>23</sup>. Fourth, we obtained data on the monthly minimum and



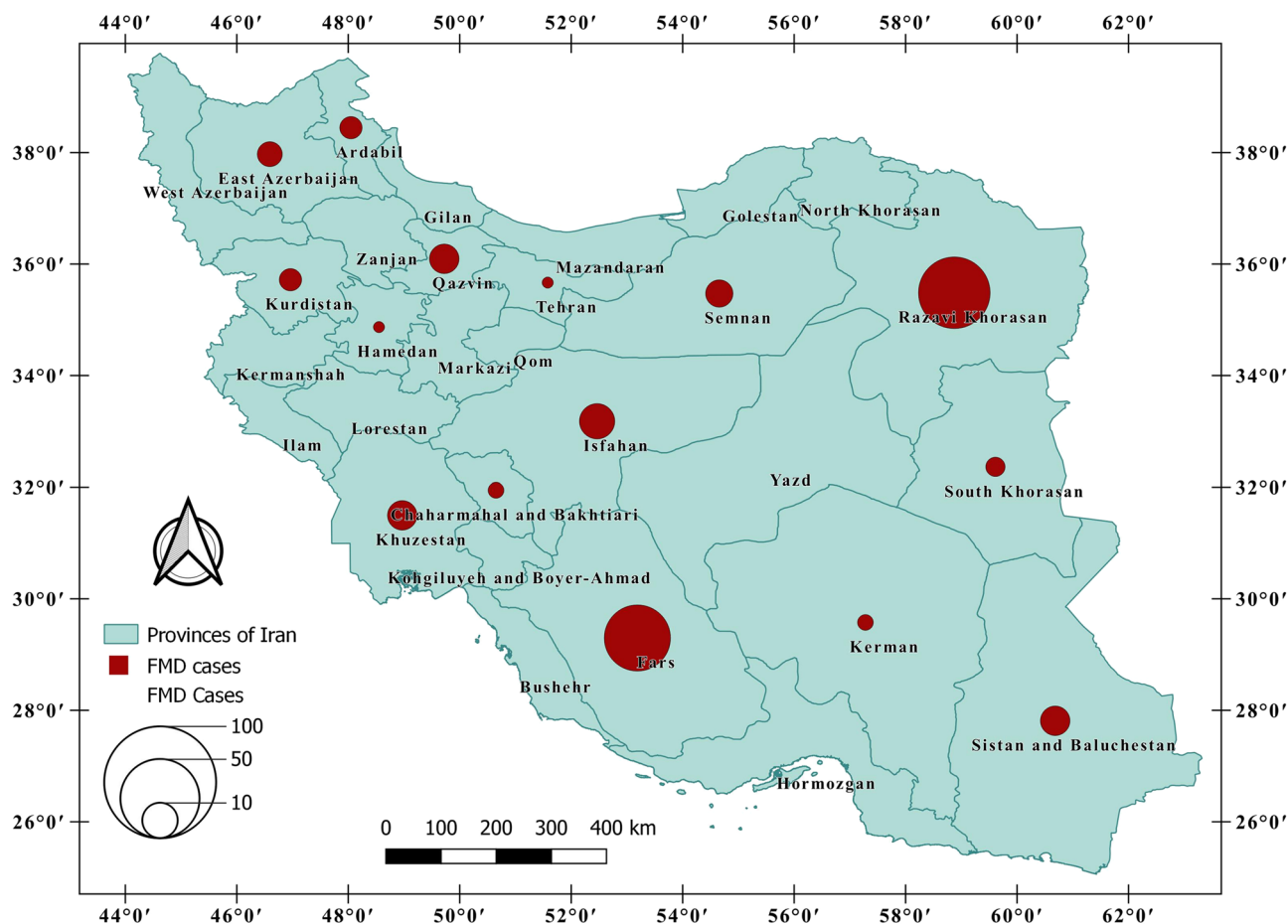
**Figure 1.** The animal population per province in Iran.

maximum temperatures for the provinces from March 2017 to March 2018, provided by the Iranian Meteorological Organization<sup>24</sup>. This data was initially in tabular format for a list of weather stations across Iran. To obtain temperature values at the provincial level, we employed the inverse distance weighted (IDW) method<sup>25</sup>. We generated interpolated temperature rasters covering the entire country using IDW. Subsequently, we calculated the average of the minimum and maximum temperatures for each month per province, utilizing the results obtained from the IDW interpolation process in the previous step. Fifth, we acquired monthly precipitation data<sup>26</sup> from the Climate Research Unit-Time Series, downsampled to a spatial resolution of 2.5 min (~21 km<sup>2</sup>) using WorldClim 2.1<sup>27</sup>. This dataset consists of 12 raster files spanning from March 2017 to March 2018, with precipitation values provided in millimeters. To derive the monthly average precipitation at the provincial level, we aggregated the raster data from March 2017 to March 2018 for each province.

Sixth, The National Geospatial-Intelligence Agency<sup>28</sup> supplied the Digital Elevation Model (DEM) with a resolution of 30 m. DEM was employed to calculate the average elevation per province. Lastly, NASA Earth Observations<sup>29</sup> provided the monthly Normalized Difference Vegetation Index (NDVI). The average NDVI value per province was derived for each 12 months from March 2017 to March 2018. The NDVI values in this study ranged from -1 to 1, where a value of 1 indicated higher vegetation. They were rescaled to a range of 0–255 and stored as unsigned 8-bit data for data storage.

**Descriptive and exploratory analysis.** First, the Pearson correlation matrix was applied to explore the correlation between all 46 explanatory variables and remove explanatory variables with low correlation to FMD incidence (dependent variables). Explanatory variables with correlation coefficients less than |0.3| (poor correlation<sup>30</sup>) were removed from the modeling process (see Supplemental Fig. 1). Then, the variance inflation factor (VIF) was employed to detect variables with a high VIF, more than 5, to remove exploratory variables with high multicollinearity<sup>31</sup>. For comparison, all global and local methods were implemented with the same chosen variables. We executed a global Moran I statistic to find the pattern of FMD incidence using ArcMap (version 10.8).

Next, Global models were executed by Python Spatial Library (PySAL) in a Python environment (version 3.7.11) using the first-order Queen contiguity to calculate the weight matrix<sup>32</sup>. Then, Local models were executed by MGWR (version 2.2) using a Fixed Gaussian kernel, set with the bandwidth and minimized corrected Akaike's Information Criterion (AIC)<sup>33</sup>. Finally, adjusted  $R^2$  and AIC were used to examine the performance of the models,

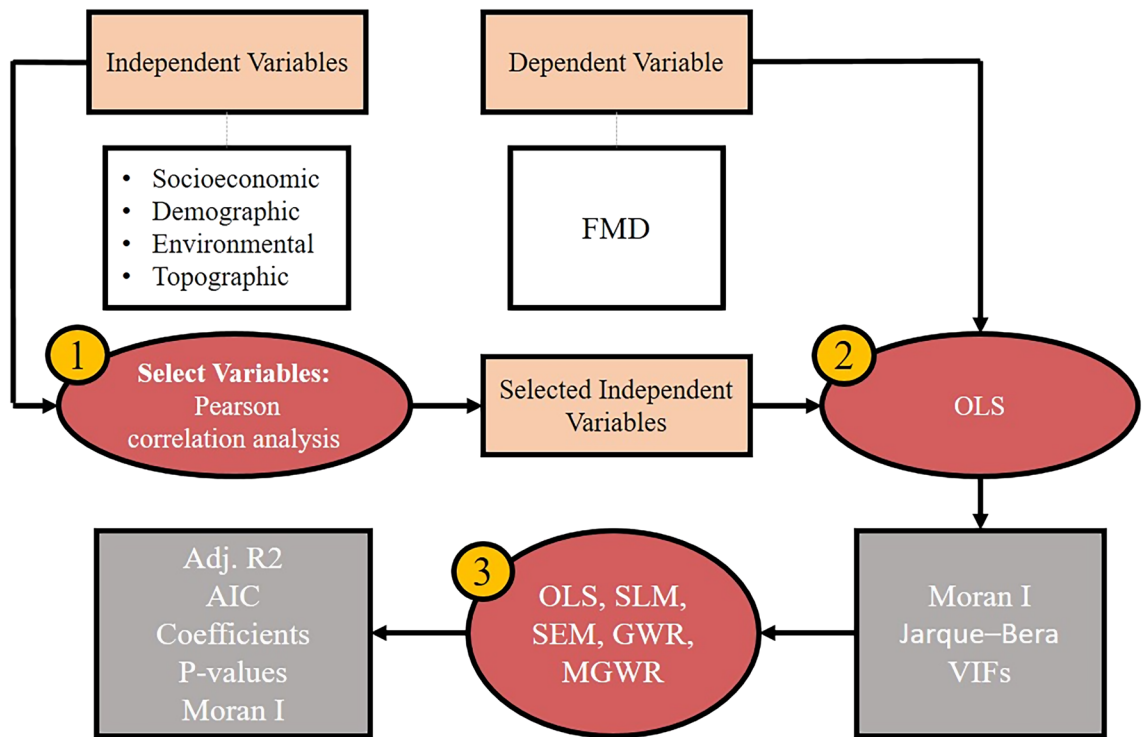


**Figure 2.** Illustrating the FMD cases from March 21, 2017, to March 21, 2018.

Theme	Variable	Description	Source
Socioeconomic	1. Unemployment rate 2. Education 3. Migration 4. GDP (without oil)	1. Unemployment rate: $\frac{\text{Number of unemployed in province}}{\text{Total population in province}} \times 100$ 2. Percentage of educated persons: $\frac{\text{Number of educated in province}}{\text{Total population in province}} \times 100$ 3. Percentage of migrated persons: $\frac{\text{Number of migrants in province}}{\text{Total population in province}} \times 100$ 4. Percentage of gross domestic product (GDP) without oil from March 21, 2017, to March 21, 2018	1. The summary of the results of the 2016 workforce statistics plan of Iran <sup>21</sup> 2, 3. Population and housing census of Iran, 2016 <sup>20</sup> 4. An overview of the gross domestic product by province from 2010 to 2018 <sup>23</sup>
Demographic	1. Population 1. Sheep 2. Goats 3. Cows	1. Percentage of population: $\frac{\text{Number of population in province}}{\text{Total population in the country}} \times 100$ 1. Percentage of sheep: $\frac{\text{Number of sheep in province}}{\text{Total sheep in the country}} \times 100$ 2. Percentage of goats: $\frac{\text{Number of goats in province}}{\text{Total goats in the country}} \times 100$ 3. The percentage of cows: $\frac{\text{Number of cows in province}}{\text{Total cows in the country}} \times 100$	1. Population and Housing Census of Iran, 2016 <sup>20</sup> 2, 3, 4. National Agricultural Census of Iran, 2014 <sup>22</sup>
Environmental	1. Temperature 1. Precipitation 2. NDVI	1. Minimum and maximum for each month (°C) (24 variables) 1. Average precipitation for each month (mm) (12 variables) 2. Normalized difference vegetation index (NDVI)	1. Monthly synoptic data of Iran from March 2017 to March 2018 <sup>24</sup> 1. CRU-TS 4.03 <sup>26</sup> downscaled with WorldClim 2.1 <sup>27</sup> 1. Obtained from NASA Earth Observation, (2022) <sup>29</sup>
Topographic	Elevation	DEM of the U.S. (30 m spatial resolution)	Obtained from National Geospatial-Intelligence Agency, (2022) <sup>28</sup>

**Table 1.** Definitions and sources of explanatory variables used in this study.

and the Moran I test for GWR and MGWR residuals were computed to test the models' residual spatial autocorrelation. Figure 3 indicates the workflow of the study.



**Figure 3.** Illustrating workflow of the study.

**Global models.** Three global regression models were applied in this study, including ordinary least square (OLS), spatial lagged model (SLM), and spatial error model (SEM). OLS is a linear aspatial regression method that can estimate the dependent variable (incidence of FMD) using a group of independent variables<sup>34</sup>. SLM is a subset of the OLS<sup>35</sup> that considers the effect of a spatial unit on adjacent units in the region, as shown in the Eq. (1)<sup>36,37</sup>. SLM explains as follows:

$$y_i = b_0 + x_i b + \rho w_i y_i + e_i \tag{1}$$

$\rho$  is the spatial auto-regressive ratio, and  $w$  is the spatial weight matrix representing distance associations between the centroid of provinces. The spatial lag function that assesses the effect of adjacent variables together can involve an independent variable during modeling<sup>36</sup>. In the SEM model,  $e_i$  is the absolute error term responsible for solving the problem of spatial autocorrelation. It splits into two components: first, the spatial component of the error term ( $\lambda w_i \xi_i$ ), and second, random error ( $e_i$ ) as follows<sup>38</sup>:

$$y_i = b_0 + x_i b + \lambda w_i \xi_i + e_i \tag{2}$$

**Local models.** Geographically weighted regression (GWR) and multiscale geographically weighted regression (MGWR) are two local spatial regression models used in this study. GWR donates to modeling spatial processes, and it can estimate the dependent variable (FMD incidence) using a group of independent variables measured at a location<sup>39</sup>. The equation of GWR is as follows:

$$y_i(u) = \beta_{0i}(u) + \beta_{1i}(u)x_{1i} + \beta_{2i}(u)x_{2i} + \dots + \beta_{mi}(u)x_{mi} \tag{3}$$

$y_i(u)$  is the dependent variable, in this case study, FMD incidence. The parameters with the notation  $x_{ji}$  are independent variables, for example, the number of sheep or precipitation in location  $i$ . In Eq. (3), we see the number of independent variables is  $m$ .  $\beta_{0i}$  shows the parameter that explains a relationship around location  $u$  and is specific to that location  $\beta^{\wedge}(u)$  takes the form as follows:

$$\beta^{\wedge}(u) = \left( X^T W(u) X \right)^{(-1)} X^T W(u) y \tag{4}$$

$W$  represents the weight matrix that considers the effect of the neighboring points relative to point  $u$ .  $X^T W(u) X$  is the geographically weighted variance-covariance matrix, and  $y$  is the vector of the values of the dependent variable (FMD incidence in this research). The leading diagonal of the  $W(u)$  matrix consists of the geographical weights, and the off-diagonal elements of the  $W(u)$  matrix are 0. The weights computed using the Gaussian kernel<sup>39</sup> as follows:

$$w_i(u) = e^{[-0.5(d_i(u)/h)^2]} \tag{5}$$

$w_i(u)$  is the geographical weight of the  $i$ th observation (centroid of the provinces) relative to the location  $u$ ,  $d_i(u)$  indicates the Euclidean distance between the  $i$ th observation and the location  $u$ , and  $h$  is a parameter named the bandwidth<sup>39</sup>. Although GWR is a significant enhancement compared to SLM and SEM regression models, the scales of relationships between the dependent and explanatory variables are assumed to be constant in GWR. Hence, MGWR is shown as follows<sup>33,40</sup>:

$$y_i = \sum_{j=0}^m \beta_{bwj} X_{ij} + \varepsilon_i, i = 1, 2, \dots, n \quad (6)$$

$\beta_{bwj}$  is the bandwidth used for calibration of the  $j$ th relationship. Practically, MGWR is a model that could be calibrated using backfitting algorithms. The MGWR can be reformulated as explained in the following equation<sup>40</sup>.

$$y_i = \sum_{j=0}^m f_{ij} + \varepsilon \quad (7)$$

$\beta_{bwj} X_{ij}$  is replaced by  $f_{ij}$ , while indicates the  $j$ th additive term and is a function utilized to  $j$ th dependent variable of  $i$ th province. Consequently, the MGWR model incorporates different bandwidths for independent variables. These varying bandwidths capture differences in spatial scales, allowing the model to capture spatial heterogeneity by considering the influence of scale on spatial processes. By incorporating different bandwidths, the MGWR model can effectively account for variations in spatial relationships across different scales, providing a more comprehensive understanding of the spatial processes<sup>33,40</sup>.

In this study, we conducted several statistical tests to assess the validity and robustness of our regression models. Firstly, we employed the Jarque–Bera test to examine the normality of the residuals obtained from the OLS regression<sup>41</sup>. The null hypothesis ( $p$  value  $< 0.05$ ) assumes that the residuals follow a normal distribution, while the alternative hypothesis ( $p$  value  $> 0.05$ ) suggests departures from normality. Secondly, we employed Moran's I test to assess the presence of spatial patterns in the distribution of FMD incidence<sup>42</sup>. Moran's I test helps us determine if there is spatial autocorrelation in the occurrence of FMD incidence, indicating whether neighboring areas exhibit similar levels of FMD incidence. Lastly, we employed the condition number (CN) to measure multicollinearity in the GWR and MGWR regression models<sup>43</sup>. The condition number quantifies the degree of multicollinearity, indicating the potential presence of highly correlated independent variables. By assessing the condition number, we can identify and address issues related to multicollinearity, which can affect the stability and interpretability of regression models.

## Results

Initially, 46 candidate independent variables were subjected to Pearson correlation analysis, which identified four variables with correlations greater than  $|0.3|$ . Specifically, these variables were the percentage of the population, precipitation in January, and the percentages of sheep and goats (see Supplemental Fig. 1). Subsequently, the ordinary least squares (OLS) model was applied to these four selected variables to assess their variance inflation factors (VIFs). As shown in Table 2, all four variables exhibited VIF values below 5, indicating no significant multicollinearity issues within the OLS model. Finally, based on these findings, the four selected variables were deemed suitable for inclusion in both the global and local spatial models, which will be further examined in the subsequent analysis.

The OLS model yielded an AIC value of 208.396. Furthermore, the Jarque–Bera statistic (value = 5.804,  $p = 0.0548$ ) indicated insufficient evidence to conclude that the OLS regression residuals deviate from a normal distribution. Based on Supplemental Table 2, Moran's I test ( $I = 0.00078$ ,  $z = 0.51011$ ,  $p = 0.60997$ ) suggests that the pattern of FMD incidence does not appear to be significantly different from random. It indicates that there is no compelling evidence of spatial clustering or dispersion in the distribution of FMD incidence.

The  $p$  values associated with the selected independent variables in Supplemental Tables 3 and 4 for the global models (OLS and SLM) indicate that the percentage of sheep and goats have  $p$  values below 0.05. It suggests that

Variable	VIFs
The percentage of population	1.14318
The percentage of sheep	1.17634
The percentage of goats	1.25663
Precipitation in January	1.17050

**Table 2.** Variance inflation factor (VIF) for the final independent variables.

Parameter/Model	OLS	SLM	SEM	GWR	MGWR
Adjusted $R^2$	0.43	0.52	0.51	0.70	0.90
AIC	208.396	209.681	208.296	57.795	27.499

**Table 3.** The value of adjusted  $R^2$  and AIC for global and local approaches in the modeling of FMD in Iran from March 21, 2017, to March 21, 2018.

these two variables are statistically significant among the independent variables and positively associated with FMD incidence per province.

Based on the results presented in Table 3, the OLS model, with the lowest adjusted  $R^2$  value of 0.43, performed the poorest among the global and local models. However, when accounting for spatial dependence, the SEM and SLM models demonstrated improved performance compared to OLS, with adjusted  $R^2$  values of 0.52 and 0.51, respectively. Moving to the analysis of local spatial differences, GWR and MGWR were utilized. Notably, the local models exhibited substantially improved adjusted  $R^2$  compared to the global models. Specifically, the adjusted  $R^2$  value for GWR was 0.70, indicating a considerable increase in explanatory power. However, MGWR outperformed all other models, achieving the highest adjusted  $R^2$  value of 0.90 and the lowest AIC value of 27.499. These results indicate that MGWR is the most effective model employed in this study, explaining 90% of the total variations in FMD incidence. Additionally, Moran's I test was conducted on the residuals of both GWR ( $I = -0.019827$ ,  $z = 0.188240$ ,  $p = 0.850688$ ) and MGWR ( $I = -0.024190$ ,  $z = 0.129640$ ,  $p = 0.896851$ ) indicated statistical insignificance, suggesting a lack of residual spatial autocorrelation in line with the model assumption.

The results of the MGWR model are presented in Supplemental Table 5. The optimal bandwidth for the independent variables ranges from 169.740 to 3618.130, indicating that these variables exhibit variation on different spatial scales. Specifically, the percentage of goats and the percentage of sheep are observed at a smaller spatial scale compared to the percentage of the population and the precipitation in January. On the other hand, the GWR model employs a fixed bandwidth of 485.560 for all variables, which does not account for the varying spatial scales of the predictors. Analyzing the spatial variability of local variables value reveals that these predictors primarily have a local impact rather than a global one, further highlighting the localized nature of their influence on FMD incidence.

Figures 4 and 5 depict the spatial variation of variables based on the GWR and MGWR models, respectively. In these figures,  $p$  values less than 0.05 in the provinces indicate statistical significance, and positive coefficients indicate a positive association between a variable and FMD incidence. Specifically, Fig. 4a represents the intercept for the GWR model, while Fig. 5a represents the intercept for the MGWR model. Figure 4b highlights a statistically significant and positive association between the percentage of the population and FMD incidence in southeastern provinces, including Sistan and Baluchestan, as well as Hormozgan provinces. It suggests that a higher population percentage in these regions is associated with an increased risk of FMD.

Figure 4c illustrates a statistically significant and positive association between the percentage of sheep and FMD incidence throughout the country, excluding the western provinces. It suggests that a higher percentage of sheep in most regions of the country is associated with an increased risk of FMD. Figure 4d displays a statistically significant and positive association between the percentage of goats and FMD incidence throughout the country, except in the eastern and northwestern provinces. It indicates that a higher percentage of goats in most regions of the country is associated with an increased risk of FMD. On the other hand, Fig. 4e reveals a statistically insignificant association between precipitation in January and FMD incidence. It suggests that precipitation in January does not significantly contribute to the variation in FMD incidence across the country. Moving on to Fig. 5b demonstrates a statistically significant and positive association between the percentage of the population and FMD incidence in specific provinces, including Busher, Hormozgan, Fars, Yazd, North Khorasan, South Khorasan, and Razavi Khorasan. However, the percentage of the population is not a statistically significant variable in other provinces. Lastly, Fig. 5c–e indicate that the percentage of sheep, goats, and precipitation in January are not statistically significant variables in Iran, according to the MGWR model.

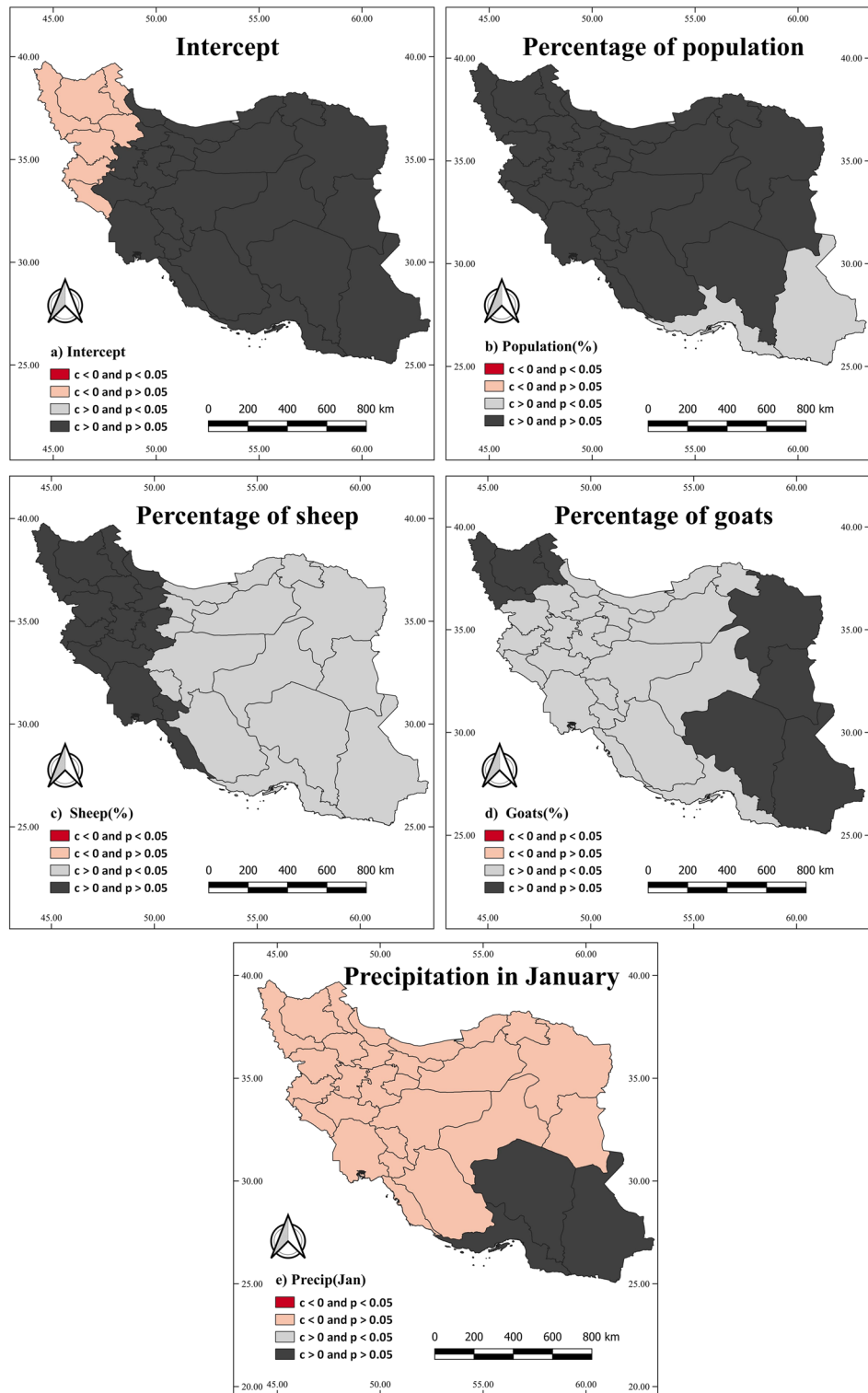
Figure 6 shows the local  $R^2$  of GWR and MGWR models used in this study. In both models, several provinces in the northeastern parts had high local  $R^2$ , which means the model performs better in these areas. In the northwest, like West Azerbaijan and East Azerbaijan provinces, the value of local  $R^2$  indicates the inadequate performance of the model.

Supplemental Fig. 3 compares the condition numbers (CN) in the GWR and MGWR models. The condition number reflects the degree of collinearity among the explanatory variables in the model. In the GWR model, some provinces in eastern Iran exhibited high condition numbers ( $CN > 3.47$ ), indicating a high degree of collinearity among the explanatory variables in those provinces. However, the MGWR model showed lower condition numbers than the GWR model, suggesting reduced collinearity among the variables in the MGWR model. It indicates that the MGWR model provides a more reliable and stable estimation of the relationships between the variables in those provinces.

## Discussion

In this study, we employed spatial modeling methods to identify the key factors contributing to FMD incidence at the province level in Iran. Out of the 46 potential variables considered, we selected four variables representing different thematic categories, including environmental, socioeconomic, demographic, and topographic factors. These variables include the percentage of the population, precipitation in January, percentage of sheep, and percentage of goats. We aimed to capture the essential determinants of FMD incidence in Iran by focusing on these variables. Our approach aligns with recent studies that have utilized spatial analysis techniques to investigate FMD patterns and drivers<sup>14,44,45</sup>. Dion and Lambin<sup>46</sup> conducted a study examining the transmission risk scenarios of FMD in southern Africa, considering climatic, social, and landscape changes, which aligns with our research. Furthermore, another study investigated the impacts of climate change, specifically abrupt temperature changes, on the risk of FMD disease in elephants across Asia and Africa, showcasing the relevance of climate-related factors in understanding FMD dynamics<sup>7</sup>.

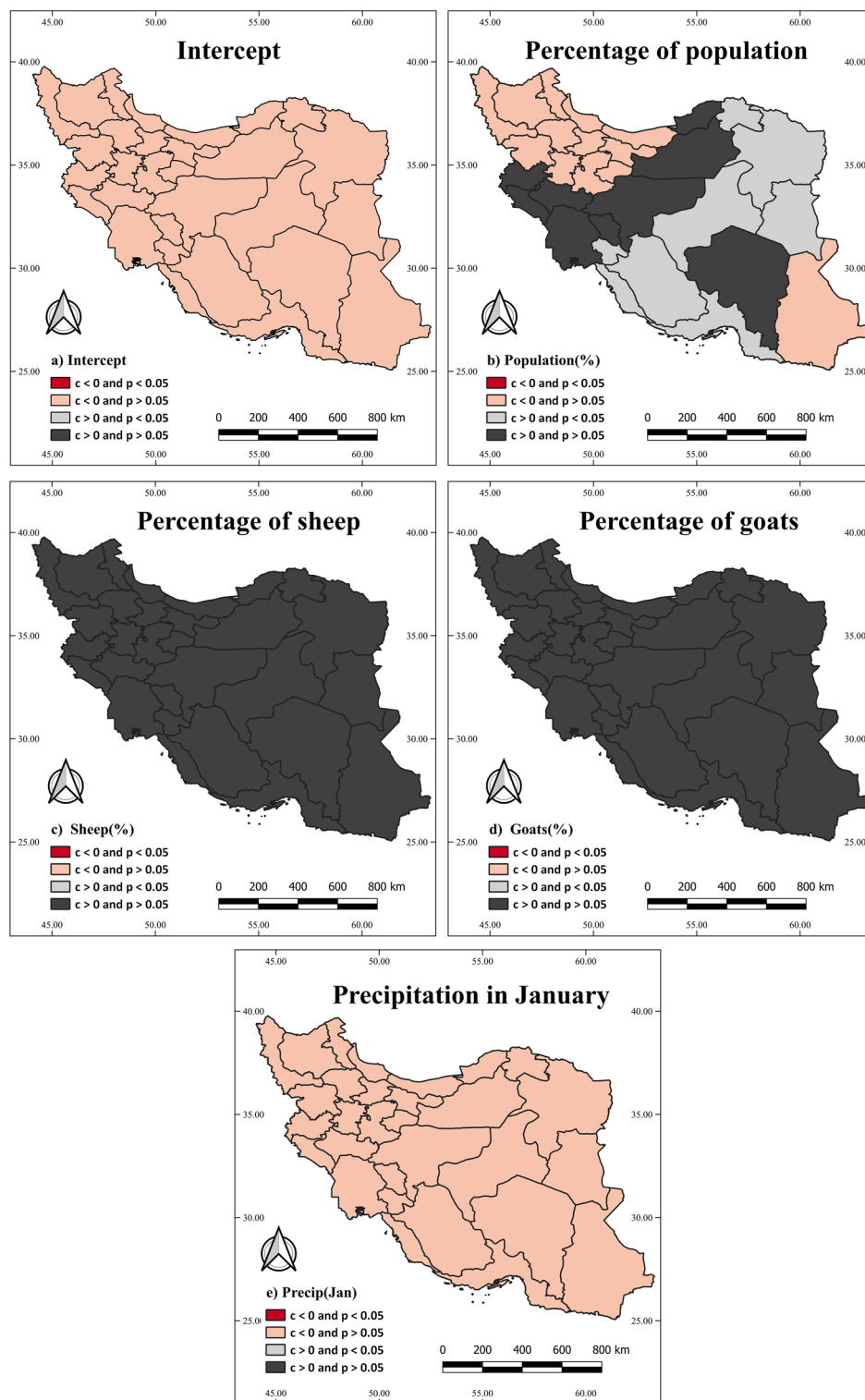
In this study, we employed a comprehensive modeling approach that included global aspatial modeling (OLS), global spatial modeling (SEM and SLM), and local spatial modeling (GWR and MGWR) to analyze the spatial distribution of FMD in Iran. Our findings highlight the superiority of the MGWR model over the traditional



**Figure 4.** Illustrating the percentage of sheep, goats, population, and precipitation in January associations with the FMD using the GWR model at the provincial level. The “c” and “p” labels indicate the coefficient and the p value, respectively.

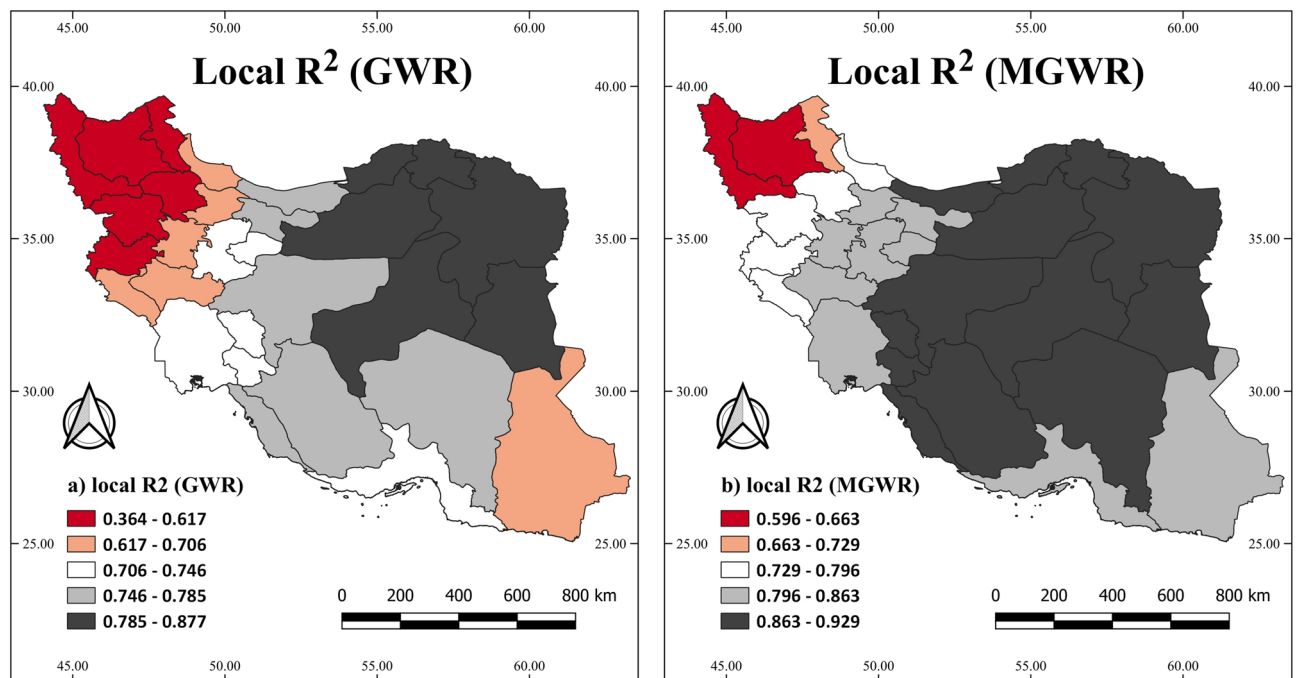
GWR model in achieving a more precise model fit. By utilizing unique bandwidths for each covariate, MGWR can capture intricate relationships that may be overlooked by the GWR model. Although this approach increases computational complexity, it provides a more nuanced understanding of spatial patterns and influences of each





**Figure 5.** Illustrating the percentage of sheep, goats, population, and precipitation in January associations with the FMD using the MGWR model at the provincial level. The “c” and “p” labels indicate the coefficient and the p value, respectively.

covariate on FMD incidence. In line with our study, previous research by Ye et al.<sup>47</sup> demonstrated a relationship between urban landscape patterns and infectious disease risk using GWR and OLS modeling techniques. Their findings indicated that GWR outperformed OLS in capturing the spatial variations of disease risk<sup>47</sup>. Zhang et al.<sup>48</sup>



**Figure 6.** The geographic distribution of the local  $R^2$  of the GWR and MGWR: the association of FMD with the percentage of sheep and goats, population, and the precipitation in January.

also investigated the spatial distribution of tuberculosis and its association with meteorological factors in mainland China. Their study revealed that GWR was a more suitable modeling approach versus OLS, as indicated by higher adjusted  $R^2$  values and lower AICc scores. The results from Ye et al. and Zhang et al. are consistent with our findings, supporting the notion that local spatial regression methods such as GWR and MGWR yield higher adjusted  $R^2$  values when compared to global regression methods (the OLS, SLM, and SEM)<sup>47,48</sup>.

Following the spatial modeling analysis, our study identified two key variables, the percentage of sheep and goats, which demonstrated a significant impact on disease incidence across most provinces in Iran. Continuous monitoring of these variables is crucial for understanding the dynamics of FMD spread at the provincial level in Iran. These findings align with the results of Begovoeva et al.<sup>49</sup> study, which identified the population of sheep, plus goats, as significant factors influencing FMD prevalence in northern Nigeria (Bauchi, Kaduna, and Plateau states). In contrast, while previous studies have highlighted the significance of climate factors such as precipitation in FMD occurrence, our study did not find precipitation to be a significant variable. This contrasts with the findings of Rahman et al.<sup>50</sup>, who investigated FMD space–time clusters and risk factors in Bangladeshi cattle and buffalo. They highlighted the substantial role of climate, particularly precipitation, in FMD incidence<sup>50</sup>. Jiang et al.<sup>7</sup> also emphasized the influence of climate change on FMD risk in elephants, noting the importance of precipitation and temperature in this context. Additionally, Lee et al.<sup>51</sup> studied the temporal patterns and space–time cluster analysis of FMD cases in Vietnam from 2007 to 2017. They identified a higher occurrence of FMD cases during the dry season, from November to March<sup>51</sup>. These variations in findings regarding the significance of precipitation in FMD incidence highlight the importance of considering regional and local factors in disease dynamics. Further research is needed to explore the specific contextual factors that influence FMD transmission patterns in different regions, considering the interplay between climatic variables, host characteristics, and local epidemiological conditions.

**Limitations.** There were several limitations to this study. Firstly, the finest spatial granularity for demographic data in Iran was available at the province level due to limited data access. However, obtaining data at the county, district, and farm levels could have provided more accurate and detailed results. Additionally, the unavailability of FMD disease data disaggregated by species restricted the analysis to the provincial level without considering the specific impact of FMD on different livestock species. Access to species-specific data would have facilitated a more comprehensive understanding of the disease dynamics and allowed for targeted prevention and control measures. Secondly, the study was limited regarding the variables used in the modeling process. While efforts were made to include relevant socio-environmental determinants, additional variables such as vaccination-related factors, as observed in previous studies, could have enhanced the analysis. For instance, considering vaccination coverage, particularly the vaccination of calves under 12 months, as an explanatory variable would have provided valuable insights into the effectiveness of vaccination strategies in controlling FMD. Thirdly, the temporal scope of the study was limited to the period from March 21, 2017, to March 21, 2018. A more extended or recent time frame would have allowed for a more comprehensive assessment of FMD incidence trends and patterns in Iran. Incorporating longer-term data could have provided insights into temporal variations and the potential impact of evolving factors on FMD incidence. Lastly, it is essential to acknowledge

that the results of this study are generalizable only at the province level. Attempting to conclude sub-province or individual levels may lead to inaccurate inferences due to the potential for ecological fallacy. Therefore, caution should be exercised when extrapolating findings beyond the analyzed spatial scale. Addressing these limitations in future studies, including obtaining data at finer spatial and temporal resolutions, incorporating species-specific FMD information, and expanding the range of variables considered, would further enhance our understanding of FMD dynamics and support more targeted and effective control strategies.

## Conclusion

In this study, we employed global and local spatial models to investigate the key factors influencing the occurrence of FMD in Iran. Our findings revealed that global models performed relatively poorly compared to local models, highlighting the importance of capturing spatial variation in FMD incidence. The MGWR model demonstrated the highest performance, with an adjusted  $R^2$  of 90% among the local models. It emphasizes the significance of considering localized spatial effects when studying FMD incidence. The results highlighted the percentage of sheep and the percentage of goats as the most significant factors among the four selected socio-environmental determinants in explaining FMD incidence across most of the provinces in Iran. It underscores the importance of considering the livestock population when making vaccination-related decisions.

## Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Received: 5 March 2023; Accepted: 17 August 2023

Published online: 19 August 2023

## References

- Mahy, B. W. Introduction and history of foot-and-mouth disease virus. *Curr. Top. Microbiol. Immunol.* **288**, 1–8. [https://doi.org/10.1007/3-540-27109-0\\_1](https://doi.org/10.1007/3-540-27109-0_1) (2005).
- Belsham, G. J., Bøtner, A., & Lohse, L. *Foot-and-Mouth Disease in Animals*. <https://www.merckvetmanual.com/generalized-conditions/foot-and-mouth-disease/foot-and-mouth-disease-in-animals> (2022).
- Thomson, G., Bastos, A., Leotta, D., Primozich, J. & Beach, K. Infectious diseases of livestock. (2004).
- Bouma, A. *et al.* The foot-and-mouth disease epidemic in The Netherlands in 2001. *Prev. Vet. Med.* **57**, 155–166. [https://doi.org/10.1016/s0167-5877\(02\)00217-9](https://doi.org/10.1016/s0167-5877(02)00217-9) (2003).
- Brito, B. P., Rodriguez, L. L., Hammond, J. M., Pinto, J. & Perez, A. M. Review of the global distribution of foot-and-mouth disease virus from 2007 to 2014. *Transbound Emerg. Dis.* **64**, 316–332. <https://doi.org/10.1111/tbed.12373> (2017).
- Aman, E., Molla, W., Gebreegizabher, Z. & Jemberu, W. T. Spatial and temporal distribution of foot and mouth disease outbreaks in Amhara region of Ethiopia in the period 1999 to 2016. *BMC Vet. Res.* **16**, 185. <https://doi.org/10.1186/s12917-020-02411-6> (2020).
- Jiang, F. *et al.* Assessing the impact of climate change on the spatio-temporal distribution of foot-and-mouth disease risk for elephants. *Glob. Ecol. Conserv.* **23**, e01176. <https://doi.org/10.1016/j.gecco.2020.e01176> (2020).
- Udahemuka, J. C. *et al.* Risk factors for the incursion, spread and persistence of the foot and mouth disease virus in Eastern Rwanda. *BMC Vet. Res.* **16**, 1–10 (2020).
- Ahmadkhani, M. & AAJAPJoTD, A. Space-time analysis of human brucellosis considering environmental factors in Iran. *Asian Pac. J. Trop. Dis.* **7**, 257–265 (2017).
- Shirzad, R., Alesheikh, A. A., Ahmadkhani, M. & Naddaf, S. R. *Aedes albopictus*: A spatial risk mapping of the mosquito using geographic information system in Iran. *Appl. Geom.* **13**, 691–700 (2021).
- Habibi, R., Alesheikh, A. A. & Bayat, S. An event-based model and a map visualization approach for spatiotemporal association relations discovery of diseases diffusion. *Sustain. Cities Soc.* **87**, 104187 (2022).
- Ballard, K. & Bone, C. Exploring spatially varying relationships between Lyme disease and land cover with geographically weighted regression. *Appl. Geogr.* **127**, 102383 (2021).
- Sangrat, W., Thanapongtharm, W. & Poolkhet, C. Identification of risk areas for foot and mouth disease in Thailand using a geographic information system-based multi-criteria decision analysis. *Prev. Vet. Med.* **185**, 105183 (2020).
- Gao, H. & Ma, J. Spatial distribution and risk areas of foot and mouth disease in mainland China. *Prev. Vet. Med.* **189**, 105311 (2021).
- Hagerman, A. D. *et al.* Temporal and geographic distribution of weather conditions favorable to airborne spread of foot-and-mouth disease in the coterminous United States. *Prev. Vet. Med.* **161**, 41–49 (2018).
- Jafarzadeh, S. R., Norris, M. & Thurmond, M. C. Prediction of province-level outbreaks of foot-and-mouth disease in Iran using a zero-inflated negative binomial model. *Prev. Vet. Med.* **115**, 101–108 (2014).
- Perez, A. M., Thurmond, M. C., Grant, P. W. & Carpenter, T. E. Use of the scan statistic on disaggregated province-based data: Foot-and-mouth disease in Iran. *Prev. Vet. Med.* **71**, 197–207 (2005).
- Ilbeigi, K., Bokaie, S., Aghasharif, S., Soares Magalhães, R. J. & Rashtibaf, M. Risk factors for recurrence of FMD outbreaks in Iran: A case–control study in a highly endemic area. *BMC Vet. Res.* **14**, 1–7 (2018).
- Noroozi, J., Akhiani, H. & Breckle, S.-W. Biodiversity and phytogeography of the alpine flora of Iran. *Biodivers. Conserv.* **17**, 493–521 (2008).
- Population and Housing Censuses. Statistical Centre of Iran. <https://www.amar.org.ir/english/Population-and-Housing-Censuses> (2016).
- Iran, T. s. o. t. r. o. t. w. s. p. o. Statistical Centre of Iran. <https://www.amar.org.ir/english/> (2016).
- National Agricultural Census of Iran, 2014. Statistical Centre of Iran. <https://www.amar.org.ir/english/> (2014).
- An overview of the gross domestic product by province during the years 2010 to 2018. Labor and Social Welfare. Iranian Centre of Statistics Ministry of Cooperatives. <https://ssicenter.mcls.gov.ir/fa/news/list/1399/8/10326>.
- Monthly synoptic data of Iran from March 2017 to March 2018. Iran Meteorological Organization (IRIMO). <https://data.irimo.ir/> (2022).
- Cao, W., Hu, J. & Yu, X. In *2009 17th International Conference on Geoinformatics*. 1–5 (IEEE).
- Harris, I., Jones, P. D., Osborn, T. J. & Lister, D. H. Updated high-resolution grids of monthly climatic observations—the CRU TS3.10 Dataset. *Int. J. Climatol.* **34**, 623–642 (2014).
- Fick, S. E. & Hijmans, R. J. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).

28. Digital Elevation Model-SRTM 1 Arc-Second 30m. USA National Geospatial-Intelligence Agency. <https://lta.cr.usgs.gov/citation> (2022).
29. Monthly NDVI. NASA Earth Observation. <https://neo.gsfc.nasa.gov/> (2022).
30. Chan, Y. Biostatistics 104: Correlational analysis. *Singapore Med. J.* **44**, 614–619 (2003).
31. O'Brien, R. M. A caution regarding rules of thumb for variance inflation factors. *Qual. Quant.* **41**, 673–690 (2007).
32. Rey, S. J. & Anselin, L. *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications* 175–193 (Springer, 2009).
33. Oshan, T. M., Li, Z., Kang, W., Wolf, L. J. & Fotheringham, A. S. mgwr: A Python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. *ISPRS Int. J. Geo Inf.* **8**, 269 (2019).
34. Zdaniuk, B. Ordinary least-squares (OLS) model. *Encyclop. Qual. Life Well-Being Res.* **20**, 4515–4517 (2014).
35. Darmofal, D. *Spatial Analysis for the Social Sciences* (Cambridge University Press, 2015).
36. Dubin, R. *Spatial and Spatiotemporal Econometrics* Vol. 18, 75–98 (Emerald Group Publishing Limited, 2004).
37. Ward, M. D. & Gleditsch, K. S. *Spatial Regression Models* Vol. 155 (Sage Publications, 2018).
38. Ghiringhelli, C., Piras, G., Arbia, G. & Mira, A. Recursive estimation of the spatial error model. *Geogr. Anal.* **55**, 90–106 (2023).
39. Charlton, M., Fotheringham, A. S. & Brunsdon, C. Geographically weighted regression white paper. *Kildare Natl. Univ. Ireland Maynooth* **20**, 1–14 (2009).
40. Fotheringham, A. S., Yang, W. & Kang, W. Multiscale geographically weighted regression (MGWR). *Ann. Am. Assoc. Geogr.* **107**, 1247–1265 (2017).
41. Jarque, C. M. & Bera, A. K. A test for normality of observations and regression residuals. *Int. Stat. Rev.* **20**, 163–172 (1987).
42. Moran, P. A. Notes on continuous stochastic phenomena. *Biometrika* **37**, 17–23 (1950).
43. Gollini, I., Lu, B., Charlton, M., Brunsdon, C. & Harris, P. GWmodel: An R package for exploring spatial heterogeneity using geographically weighted models. [arXiv:1306.0413](https://arxiv.org/abs/1306.0413) (arXiv preprint) (2013).
44. Chanchaidechachai, T., de Jong, M. C. & Fischer, E. A. Spatial model of foot-and-mouth disease outbreak in an endemic area of Thailand. *Prev. Vet. Med.* **195**, 105468 (2021).
45. Munsey, A. *et al.* Spatial distribution and risk factors for foot and mouth disease virus in Uganda: Opportunities for strategic surveillance. *Prev. Vet. Med.* **171**, 104766 (2019).
46. Dion, E. & Lambin, E. F. Scenarios of transmission risk of foot-and-mouth with climatic, social and landscape changes in southern Africa. *Appl. Geogr.* **35**, 32–42 (2012).
47. Ye, Y. & Qiu, H. Using urban landscape pattern to understand and evaluate infectious disease risk. *Urban For. Urban Green.* **62**, 127126 (2021).
48. Zhang, Y. *et al.* Spatial distribution of tuberculosis and its association with meteorological factors in mainland China. *BMC Infect. Dis.* **19**, 1–7 (2019).
49. Begovoova, M. *et al.* Factors associated with foot-and-mouth disease seroprevalence in small ruminants and identification of hot-spot areas in northern Nigeria. *Prev. Vet. Med.* **212**, 105842 (2023).
50. Rahman, A. A. *et al.* Foot-and-mouth disease space-time clusters and risk factors in cattle and buffalo in Bangladesh. *Pathogens* **9**, 423 (2020).
51. Lee, H. S., Pham, T. L. & Wieland, B. Temporal patterns and space-time cluster analysis of foot-and-mouth disease (FMD) cases from 2007 to 2017 in Vietnam. *Transbound. Emerg. Dis.* **67**, 584–591 (2020).

## Acknowledgements

We thank the editor and the anonymous reviewers for their excellent suggestions to improve the original draft of the manuscript.

## Author contributions

M.N.A., A.A.A., and A.L. contributed to the study design. All authors (M.N.A., A.A.A., Z.N.S., A.L., S.B., S.A., and B.H.) contributed to data gathering and interpretation of the results. M.N.A. wrote the first draft of the manuscript. A.L. supervised and validated the study. All authors (M.N.A., A.A.A., Z.N.S., A.L., S.B., S.A., and B.H.) read, commented, and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-40865-4>.

**Correspondence** and requests for materials should be addressed to B.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023