Using an Eigenvector Spatial Filtering Based Spatially Varying Coefficient Model to Analyze the Spatial Heterogeneity of COVID-19 and Its Influencing Factors in Mainland China

Meijie Chen

Wuhan University

Yumin Chen (✉ ymchen@whu.edu.cn)

Wuhan University <https://orcid.org/0000-0001-5045-6418>

John P Wilson

University of Southern California

Huangyuan Tan

Wuhan University

Tianyou Chu

Wuhan University

Research Article

Keywords: COVID-19, spatial heterogeneity, Eigenvector Spatial Filtering, spatially varying coefficients, GWR

Posted Date: September 27th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-910492/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at ISPRS International Journal of Geo-Information on January 15th, 2022. See the published version at <https://doi.org/10.3390/ijgi11010067>.

1 Using an Eigenvector Spatial Filtering based spatially varying 2 coefficient model to analyze the spatial heterogeneity of COVID- 3 19 and its influencing factors in mainland China

4 Meijie Chen^a, Yumin Chen^{a,*}, J.P. Wilson^b, Huangyuan Tan^a, Tianyou Chu^a

5 ^a School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079,
6 China;

7 Chen_meijie@whu.edu.cn; tanhuangyuan@whu.edu.cn; chutianyou@whu.edu.cn

8 ^b Spatial Sciences Institute, University of Southern California, Los Angeles, CA, 90089-
9 0374, USA; jpwilson@usc.edu

10 * Correspondence: ymchen@whu.edu.cn; Tel.: +86-27-6877-8386

11

12 Abstract

13 ● **Background:** The COVID-19 pandemic has led to many deaths and economic disruptions
14 across the world. Several studies have examined the effect of health risk factors on COVID-19
15 rates in different places, but the problem of spatial heterogeneity has not been adequately
16 addressed.

17 ● **Methods:** In this paper, we developed an Eigenvector Spatial Filtering based spatially
18 varying coefficient model (ESF-SVC) to reveal the spatially varying impact of certain health
19 risk factors on the COVID-19 spread. The experiment was conducted during 7 weeks within
20 two study extents (Hubei province and mainland China). Spatial varying coefficient maps were
21 produced for spatial pattern discovery.

22 ● **Results:** Results showed that the ESF-SVC model could take good control of over-fitting
23 problems, with average adjusted R^2 16.31% (in Hubei province) and 10.25% (in mainland
24 China) higher than that of GWR. The cross validation RMSE of ESF-SVC model was also the

lowest. In Hubei province, Population density and wind speed had a significant impact on COVID-19 infection rates and that their effect was constant across cities. While in mainland China, migration score, building density, temperature and altitude showed significant impact and their effect varies across space. The influence of migration score was less contributive and less significant in cities around Wuhan than cities farther away, while the altitude showed stronger contributions in high altitude cities.

- **Conclusions:** Our study hopes to provide not only a feasible path to solve the problem of spatial autocorrelation and spatial heterogeneity in COVID-19 characterization but also an intuitive way to discover spatial patterns in large study areas, which could help people and government awareness of the potential health risks and shed some light in COVID-19 control strategies.

Keywords: COVID-19; spatial heterogeneity; Eigenvector Spatial Filtering; spatially varying coefficients; GWR

1. Introduction

The COVID-19 pandemic, initially found in Wuhan, China in December 2019, is incredibly infectious and has caused large impacts across the world (He et al., 2020). The World Health Organization (WHO) designated the COVID-19 as a global pandemic on March 11, 2020 (Cucinotta and Vanelli, 2020). The virus has spread rapidly worldwide and infected cases were found in almost all countries by the end of 2020, with global cumulative cases above 79 million and the death cases above 1.7 million, and the number kept rising (WHO, 2020).

Given this background, researchers have conducted a large number and variety of studies. Many have focused on trend analysis and time-series prediction. The susceptible-exposed-infected-removed model (SEIR) and its expansions have been used in multiple studies (Alzahrani et al., 2021; Chadsuthi and Modchang, 2021; Fanelli

and Piazza, 2020; Liu et al., 2020; Reiner et al., 2021; Yang et al., 2020) while other researchers have used machine learning methods (ArunKumar et al., 2021; Ceylan, 2020; Desai, 2021; Malki et al., 2020; Niazkar and Niazkar, 2020). The studies that focus on trend prediction could not only effectively estimate the outbreak point and turning point of the COVID-19 pandemic but also help to evaluate the effectiveness of measures and whether the strategies should be strengthened(Chang et al., 2021; EL Azzaoui et al., 2021; Guleryuz, 2021; Khedher et al., 2021; Liao et al., 2020; Saba and Elsheikh, 2020). However, these studies mostly did not consider the influence of the external risk factors, which was also important in epidemic analysis(Jia et al., 2019).

Many researchers have explored the relationship between socioeconomic as well as infrastructural factors and the spread of the COVID-19 virus (Damialis et al., 2021; Dellicour et al., 2021; Hawkins et al., 2020; Kim et al., 2021; Ma et al., 2020; Martins et al., 2020; Ravindra et al., 2021; Scarpone et al., 2020; P. Shi et al., 2020). Some found that the COVID-19 had a more significant impact on poor areas (Kulu and Dorey, 2021; Messner, 2020; Rahman et al., 2020) while a few found the influence of income was not significant in certain study areas(Andersen et al., 2021; Kwok et al., 2021; Mateo-Urdiales et al., 2021). Other researchers assessed the relationship between demographic factors and COVID-19 infection and results showed that area with higher percentage of elderly people was associated with higher infection rate(Dowd et al., 2020; Tian et al., 2020). Factors as population density, population mobility and the accessibility to hospitals were also considered in relation to the spread of COVID-19 in many studies (Li et al., 2020; Rocklöv and Sjödin, 2020; Sirkeci and Murat Yüceşahin, 2020; Sun et al., 2020; Yin et al., 2021). There are also a number of researchers estimated the association between environmental factors and the COVID-19 transmission. The influence of air quality, wind, humidity as well as temperature on COVID-19 cases were

75 mainly discussed (Bashir et al., 2020; Gupta et al., 2020; Lim et al., 2021; Qi et al.,
76 2020; Şahin, 2020; Zhu et al., 2020). Prata et al. (2020) applied a generalized additive
77 model to explore the effect of annual average temperature in Brazil and found that the
78 rise of average temperature lower than 25.8°C could decrease cumulative confirmed
79 cases of COVID-19. Zhou et al. (2020) found that the increase of median altitude may
80 result in a decrease of COVID-19 mortality rates, as people in high altitude areas had a
81 better tolerance to hypoxia. The studies about the influence of external factors on
82 COVID-19 spread could help people understand potential risks and take appropriate
83 anti-pandemic measures. However, in these studies, the spatial relationships among
84 cities or countries were underestimated (W. Shi et al., 2020). As the spread of infectious
85 diseases is often connected with spatial patterns (Elliott and Wartenberg, 2004; Jia et
86 al., 2020, 2019), that is, spatially autocorrelated (Tobler, 1970), traditional statistical
87 methods are not so suitable because the data violates the independence criterion.

88 Therefore, to solve this problem, spatial models such as spatial lag models (SLMs)
89 and spatial error models (SEMs) were utilized (Andersen et al., 2021; Kim et al., 2021;
90 Sannigrahi et al., 2020). These studies considered spatial autocorrelation from a global
91 point of view. However, one problem is that the spatial epidemic data often exhibit high
92 spatial heterogeneity (Beale et al., 2008; Elliott and Wartenberg, 2004). Global models
93 are not helpful for examining the effect of risk factors that may vary across space. Some
94 researchers use geographical weighted regression (GWR) models and its variations
95 (Brunsdon et al., 1996; Fotheringham et al., 2017) to consider spatial varying
96 characteristics (Han et al., 2021; Karaye and Horney, 2020; Maiti et al., 2021; Mansour
97 et al., 2021; Mollalo et al., 2020; Snyder and Parks, 2020; Wu et al., 2020). Although
98 GWR models can consider the effect of risk factors in a local way and spatial non-
99 stationarity, there is often multi-collinearity bias in the coefficients as well as the

problem of choosing the appropriate bandwidth (Wheeler and Tiefelsdorf, 2005). Griffith (2008) developed an eigenvector spatial filtering based spatially varying coefficients (ESF-SVC) method to take control of spatial heterogeneity. One benefit of this method is that it decomposes the spatial weights matrix into eigenvectors which contain information of different spatial patterns (Griffith, 2008). Murakami et al. (2017) developed a ESF-SVC modeling approach by considering random effects and found the ESF-SVC method was better than the GWR approach in terms of model accuracy (Murakami et al., 2017; Tan et al., 2020). Another advantage of the ESF-SVC method is that it can detect whether the coefficients of independent variables need to be spatially varying or constant through the Bayesian information minimization (BIC) criteria, and the output of the ESC-SVC model includes the estimated coefficients (be it spatially varying or constant) as well as their statistical significance (Murakami, 2017). However, this method has not been used in the study of COVID-19.

Encouraged by the aforementioned studies, we construct an ESF-SVC model to reveal the spatially varying impact of certain socio and environmental factors on COVID-19 spread. The main objectives of this paper are: (1) to explore how selected health risk factors are related the COVID-19 infection rate within different study extents; (2) to find out if the influence of selected health risk factors vary across space and time and how they vary. Considering data availability and reasonability, socio factors including population density, human migration, hospital capacity, GDP and environmental factors including precipitation, wind speed, temperature, average altitude and building density were used as the initial health risk factors according to literature reviews mentioned above. The experiment was conducted over 7 weeks from January 22, 2020 to March 10, 2020 within two study extents (Hubei province and mainland China). Spatial varying coefficient maps were used for spatial pattern

discovery and how the effect of health risk factors vary across cities. The ESF-SVC model results were compared with those of the OLS, ESF and GWR models. Our study hopes to provide not only a feasible path to solve the problem of spatial autocorrelation and spatial heterogeneity in COVID-19 studies but also an intuitive way to discover spatial and temporal patterns.

The rest of this paper is organized as follows. Section 2 introduces the study area, data resources and the corresponding pre-processing approaches as well as the proposed method. Section 3 presents the results of spatial modeling within two study extents (Hubei province and mainland China). Section 4 was the discussion about model performance, the influence of selected health risk factors as well as limitations, while Section 5 was the conclusion of this paper and ideas of future study.

2. Data and methods

2.1 Study area

Hubei province as well as mainland China are taken as study areas and the experiments were conducted at the city level (Figure 1). Considering data quality and availability, 17 and 362 cities were selected in Hubei province and mainland China (excluding Hong Kong, Macao, Taiwan, Zhoushan and Sansha), respectively.

Figure 1 Two study areas (mainland China and Hubei province)

2.2 Data resources and pre-processing

The daily confirmed cases of COVID-19 from January to April were collected from the Tencent news website. Daily climate data including average temperature (TEMP), precipitation (PRCP), wind speed (WDSP), air pressure (PRE) as well as altitude (DEM) were taken as natural influencing factors. The socioeconomic data such as Gross Domestic Product (GDP) and population data were collected from the Provincial Statistical Year Book of 2018. The daily outflow migration data in Wuhan from December 31, 2019 to March 30, 2020, which represent the scale and proportion of people who left Wuhan to other cities, was harvested from Baidu qianxi map. The data were applied to both two study scopes. The data sources that were used are listed in Table 1.

Table 1. Data resources

Name	Data resources
COVID-19 infections	https://news.qq.com/zt2020/page/feiyang.htm

TEMP, PRCP, WDSP, PRE	https://www7.ncdc.noaa.gov/CDO/cdo
DEM	http://www.resdc.cn/data.aspx
Population, GDP	statistical year book of 2018
	https://data.cnki.net/Yearbook
Outflow migration data	https://qianxi.baidu.com/
Hospitals and other buildings	Baidu map API

To explore the relationship between COVID-19 infection rates and health risk factors, the weekly average COVID-19 infection rate (denoted as WAVG) was taken as the dependent variable and a power transformation was used (Box and Cox, 1964) to make the data nearly normally distributed:

$$WAVGi = \left(\frac{1}{7} \sum_{k=1}^7 \frac{confirmed_k}{pop} \right)^{pow} \quad (1)$$

where $confirmed_k$ was the number of confirmed cases on the k th day of week i , pow was the corresponding “Box-Cox” power transformation parameter and pop was the number of residents. Population density (PDEN) was calculated by population/area(km²).

As the incubation period of COVID-19 cases generally ranged from 0 to 14 days, for the consideration of time lag effects, meteorological variables in the previous two weeks (week $i-2$ and week $i-1$) were averaged in the modeling of average COVID-19 infection rate in week i (Ma et al., 2021; Runkle et al., 2020). The meteorological data covered 376 stations with daily records. TEMP, PRCP and WDSP were averaged by every two weeks and interpolated to cover the whole study area.

The migration data obtained from Baidu qianxi map include two parts: the migration scale and migration proportion. The migration scale vector $M1$ ($1*m$) represents the total number of people that moves out from Wuhan to other places within m days, while

the daily migration proportion (M2) is a $m*n$ vector which shows the daily percentage of Wuhan migrants who moved to the other n cities within m days. The migration score vector MS ($1*n$) was calculated by $M1*M2$.

The infrastructure data including location and type of hospitals as well as other buildings was obtained with the Baidu map API. Other buildings such as shopping malls, hotels, restaurants, and tourist attractions also contribute to population gatherings and the spread of the epidemic, so the density of these types of buildings (BD) in each city was calculated and selected as a risk factor. The hospital capacity (HOS) was calculated by a weighted sum of hospitals at different level within a city:

$$HOS = 0.5 * hos3A + 0.25 * hos3 + 0.15 * hos2 + 0.1 * hos1 \quad (2)$$

where *hos3A* was the number of tertiary level-A hospitals, *hos3* was the number of other tertiary hospitals. As tertiary hospitals, especially 3A hospitals, contain the most important medical resources in China, they had a relatively higher weight(Chen et al., 2018). *hos2* was the number of secondary hospitals and *hos1* was the number of other hospitals.

All 10 risk factors mentioned above were normalized and standardized for better comparison. Then Pearson correlation tests and multicollinearity diagnosis using variance inflation factor (VIF) were applied before modeling. If the VIF value was smaller than 10, the multicollinearity within variables was not serious, otherwise variables should be excluded by stepwise regression method.

2.3 Methods

The experiment was conducted at two spatial extents over 7 weeks from January 22, 2020 to March 10, 2020. The methodology included 5 main tasks: (1) spatial weights matrix construction; (2) eigenvector extraction; (3) variable selection and ESF-SVC

199 model construction; (4) model assessment and comparison; and (5) spatial pattern
200 discovery.

201 2.3.1 Spatial weights matrix construction

202 The spatial weights matrix W we use in the study is a K nearest neighbors' contiguity
203 matrix, which was a distance-based matrix and could deal with the problem of "isolation"
204 units (Gerkman and Ahlgren, 2014; Lesage and Fischer, 2008; Schneider et al., 2021).
205 For each study unit i , calculate the distance between the centroid of unit i and the other
206 $n-1$ units ($d_{i1}, d_{i2} \dots d_{i-1}, d_{i+1}, \dots, d_n$) and rank by distance. Then find the k closest
207 units to i and denoted these as neighbors, while other units are not. When cities i and j
208 are neighbors, W_{ij} takes on non-zero values (one, for a binary matrix), otherwise zero
209 (Rogerson, 2001).

210 Then the spatial weights matrix W is centered as below:

$$211 \quad C = \left(I - \frac{11^T}{n} \right) W \left(I - \frac{11^T}{n} \right) \quad (2)$$

212 where n represents the city polygon units in this study, $n=17$ in the study of Hubei
213 province and $n=362$ in the study of mainland China, I is an n -dimension identity matrix
214 and 1 is an n -by-1 vector of ones.

215 2.3.2 Eigenvector extraction

216 The centered spatial weight matrix C in Equation 2 was decomposed into
217 eigenvectors and eigenvalues as follows:

$$218 \quad C = E \Lambda E^T \quad (3)$$

219 where $E = (E_1, E_2, E_3, \dots, E_n)$ is the collection of eigenvectors; Λ is an $n \times n$ diagonal
220 matrix of eigenvalues ($\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n$) which defines the corresponding eigenvectors'

221 MC value, λ_1 is the largest and λ_n is the smallest eigenvalue.

222 To capture the positive spatial autocorrelation, the subset of eigenvectors C was first
 223 selected using the criterion $\frac{\lambda_i}{\lambda_1} > 0.25$ for the following model. Then a stepwise variable
 224 selection method was applied in further eigenvector extraction to avoid
 225 multicollinearity(Helbich and Griffith, 2016).

226 2.3.3 Variable selection and ESF-SVC model construction

227 ESF regression uses a set of eigenvectors as new variables. The vectors extracted in
 228 Section 2.2.2 are added as control variables to the model, which is expressed as follows:

$$229 \quad Y = X\beta + E\alpha + \varepsilon \quad (4)$$

230 where Y represents the dependent variable and X represents an $n \times p$ matrix whose p -
 231 th row of element is the independent variable p , E represents an $n \times k$ matrix of k
 232 selected eigenvectors, and ε represents disturbances.

233 An extended ESF-based SVC model could be shown as:

$$234 \quad \hat{Y} \approx \left(\beta_0 \mathbf{1} + \sum_{k_0=1}^{K_0} \mathbf{E}_{k_0} \beta_{k_0} \right) + \sum_{p=1}^P \left(\beta_p \mathbf{1} + \sum_{k_p=1}^{K_p} \mathbf{E}_{k_p} \beta_{k_p} \right) \cdot \mathbf{X}_p + \varepsilon \quad (5)$$

235 where $\mathbf{X}_p$ is a $n \times 1$ vector of the p -th health risk factor, $\mathbf{E}_{k_p}$ is the k_p -th EV that
 236 combined with its corresponding health risk factor p , $\beta_0, \beta_{k_0}, \beta_{k_p}$ are the regression
 237 coefficients that estimated by restricted maximum likelihood method, ε represents
 238 random disturbance, and " $\cdot$ " denotes the element-wise product operator.

239 For modeling COVID-19 cases, Equation (5) can be expressed as:

$$240 \quad \begin{aligned} COVID - 19 IFR = & \beta_0^{R-ESF} + \beta_1^{R-ESF} PDEN + \beta_2^{R-ESF} GDP + \beta_3^{R-ESF} MS + \\ & \beta_4^{R-ESF} HOS + \beta_5^{R-ESF} BD + \beta_6^{R-ESF} TEMP + \beta_7^{R-ESF} PRCP \\ & + \beta_8^{R-ESF} PRE + \beta_9^{R-ESF} WDSP + \beta_{10}^{R-ESF} DEM \end{aligned} \quad (6)$$

241 where

$$\beta_k^{R-ESF} = \beta_{k0} + \sum_{k=1}^k E\gamma_k$$

The coefficients β_k^{R-ESF} include two parts: The first part β_{k0} represents the global effect of health risk factors on the COVID-19 infection rate and the second part $E\gamma_k$ represents the spatial effect on health risk factors and their local influence on the COVID-19 infection rate.

2.3.4 Model assessment and comparison

Four models with Gaussian link function- OLS, GWR, ESF and ESF-SVC - were built and compared using the criteria specified below.

Firstly, the adjusted R^2 , the AIC, the RMSE was selected as performance criteria to assess model reliability, calculated as follows:

$$\text{adj. } R^2 = \sqrt{1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 / \text{df}}{\sum_{i=1}^n (Y_i - \bar{Y})^2 / (n - 1)}} \quad (7)$$

$$\text{RMSE} = \sqrt{\sum_{i=1}^n (Y - \hat{Y})^2 / n} \quad (8)$$

where Y_i is the COVID-19 infection rate of the i th city, $\hat{Y}_i$ is the predicted model value of the i th city, $\bar{Y}$ is the total average COVID-19 infection rate, and df is the degrees of freedom of the corresponding model.

Then cross validation method was utilized to further test model stability and fitting accuracy. Models with better cross validation performance could be viewed as bearing less over-fitting problem (Ghojogh and Crowley, 2019; Picard and Cook, 1984). Instead of 10-fold cross validation, a more time consuming but unbiased leave-one-out cross validation method (LOOCV) was applied to alleviate the problem of spatial autocorrelation that might exist within the test dataset (Brenning, 2012; Refaailzadeh et

al., 2009). That is, only one city with its corresponding variables acts as the validation sample and the other cities serve as the training set. The process was repeated until each city were taken as a validation sample and the average RMSE was taken as the assessment criteria.

Finally, the Moran's I value for residuals was also taken as a performance criterion to test if spatial autocorrelation in the residuals was successfully filtered out. If the Moran's I value is high and significant, it suggests that the residuals are spatially autocorrelated, which violates the hypothesis of independence in linear regression.

Generally, the higher the adjusted R^2 , the lower the RMSE, AIC and the RMSE of LOOCV, the smaller and less significant the Moran's I for residuals, the better the model performance.

2.3.5 Spatial pattern discovery

The output of the ESF-SVC model contains the coefficients (be they spatially varying or constant) as well as their statistical significance, indicating how and to what extent risk factors affect the COVID-19 infection rate and how they vary across space. For better visualization, we mapped the coefficients of corresponding risk factors through 7 weeks. Quantile method was used to grade the legend and each risk factor shared the same mapping dimension. Cities whose coefficient significance was less than 0.1 were labeled by scattered points. A total of 28 (4 coefficients * 7 weeks) coefficient maps were displayed. By comparing the coefficients of different risk factors across space in 7 weeks, we can explore how the effect on the COVID-19 infection rate varies in different cities over time. For better visualization, flowchart of the research method and experimental procedures was shown in Figure 2.

Figure 2. Flowchart of research method and experimental procedures

3. Results

3.1 Hubei province

3.1.1 Descriptive statistics

Table 2 was the descriptive statistics for average health risk factors and COVID-19 infection rate during the study period in Hubei province. Variables were standardized and normalized for better comparison in modeling. Table 3 was the Moran's I value for COVID-19 infection rate through 7 weeks. The Moran's I value range from 0.342 to 0.450 at a significant level of 0.001, indicating that the COVID-19 infection rate within

296 Hubei province was highly spatially autocorrelated.

297 Table 2. Summary of the dataset in Hubei province

Value	COVID-19	PDEN	HOS	MS	GDP	BD	PRCP	TEMP	WDSP	PRE	DEM
	(case/10000 person until March 10)	(person/ km ²)	(count / km ²)			(count/ km ²)	(inch)	(°C)	(knots)	(hPa)	(m)
Min	8.908	14.96	0.30	0.00	20.95	0.01	0.02	42.46	2.49	161.16	26.86
Max	433.309	997.56	27.60	12.68	10905.60	0.39	6.23	50.06	5.83	961.29	1686.69
Mean	66.020	278.52	4.71	1.06	1797.59	0.06	1.23	47.35	4.28	390.57	327.34
Std.dev	99.250	223.48	6.33	3.01	2528.61	0.09	1.66	2.32	0.80	306.37	461.56

298

299 Table 3. Moran's I for COVID-19 infection rate in Hubei province

Time period	Moran's I	P-value
Week1	0.342	<0.001
Week2	0.385	<0.001
Week3	0.433	<0.001
Week4	0.450	<0.001
Week5	0.446	<0.001
Week6	0.443	<0.001
Week7	0.440	<0.001

300

301 3.1.2 Correlation analysis and multicollinearity diagnosis

302 Table 4 shows the Pearson correlation coefficients between health risk factors and
303 COVID-19 infection rate in Hubei province. In the first 3 weeks from January 22 to
304 February 11, all 10 risk factors did not show significant correlation with the COVID-
305 19 infection rate. This might be because the nucleic acid detection method was not
306 mature and detection resources was not sufficient at first, such that many patients were
307 not detected as infected very quickly in high risk areas (i.e., Hubei province). The
308 Chinese Healthcare Commission announced an improved method for detecting
309 confirmed cases on February 12, which was the start time point of week 4 in our study.
310 Therefore, from week 4 to week 7, PDEN, MS, BD and WDSP began to show
311 significant positive correlation while PRE and DEM showed significant negative

correlation. PRCP passed the significance test from week 3 to 4. Other risk factors including HOS, GDP, PRCP and TEMP did not pass the significance test. The VIF was shown in Table 5 to test if multicollinearity exists. The VIF of PDEN, HOS, MS, GDP and BD was larger than 10, which means multicollinearity problem exists. Therefore, a stepwise regression method was applied for variable selection to avoid multicollinearity. Finally, only PDEN and WDSP was maintained and other variables was screened out. Table 6 shows the VIF value for factors after stepwise selection and the values are smaller than 10, which means the remaining variables were not multicollinear and could be used for modeling.

Table 4. Pearson correlation coefficients between WAVG and risk factors in Hubei province

Time period	PDEN	HOS	MS	GDP	BD	PRCP	TEMP	WDSP	PRE	DEM
Week1	0.590	0.636	0.521	0.598	0.593	-0.075	0.513	0.356	0.453	-0.348
Week2	0.529	0.503	0.589	0.517	0.600	0.099	0.226	0.57	-0.383	-0.339
Week3	0.526	0.535	0.510	0.554	0.646	0.752	-0.058	0.475	-0.512	-0.516
Week4	0.847***	0.608	0.700**	0.618	0.753***	0.754	0.02	0.587*	-0.567*	-0.607*
Week5	0.865***	0.601	0.711**	0.621	0.770***	0.462	-0.204	0.548**	-0.594**	-0.609**
Week6	0.867***	0.614	0.715**	0.625	0.774***	0.465	0.375	0.666**	-0.604**	-0.608**
Week7	0.869***	0.618	0.719**	0.629	0.778***	0.293	0.331	0.674**	-0.607**	-0.606**

* Significant at 0.1 level; ** Significant at 0.05 level; *** Significant at 0.01 level

Table 5. Variance Inflation Factor (VIF) of exploratory variables in Hubei province

Time period	PDEN	HOS	MS	GDP	BD	PRCP	TEMP	WDSP	PRE	DEM
Week1	55.270	29.969	101.467	30.543	185.873	1.876	14.756	11.737	18.651	7.879
Week2	39.863	32.478	147.708	21.589	172.429	1.864	4.246	3.970	11.681	20.239
Week3	43.940	28.406	152.646	13.499	198.129	6.236	1.743	3.125	9.395	10.919
Week4	45.979	31.843	155.576	16.063	213.393	6.254	1.919	2.632	10.383	12.529
Week5	44.517	25.721	121.521	17.097	190.554	6.275	1.876	3.350	9.033	10.554
Week6	56.205	25.420	138.165	15.531	223.637	6.681	4.008	3.735	11.868	9.383
Week7	40.733	26.986	134.071	21.618	197.476	3.823	5.437	2.491	15.059	11.191

Table 6. Variance Inflation Factor (VIF) for exploratory variables after stepwise selection

Time period	PDEN	WDSP
Week1	1.000	1.000
Week2	1.332	1.332
Week3	1.481	1.481

Week4	1.103	1.103
Week5	1.169	1.169
Week6	1.287	1.287
Week7	1.092	1.092

326

327 3.1.3 Model assessment

328 The results of different models (OLS, ESF, GWR, ESF-SVC) are shown in Table 7.

329 In terms of model fitness, the ESF-SVC model performed the best, with an average
330 adjusted R^2 of 0.770, higher than that of GWR (0.662), ESF (0.730), and OLS (0.648).

331 The ESF-SVC model's average RMSE is the smallest (0.310), but the AIC is larger than
332 the other models, because the spatially varying coefficients increase model complexity.

333

Table 7. Model performance comparison in Hubei province

	OLS			ESF			GWR			ESF-SVC		
	Adj.R ²	AIC	RMSE	Adj.R ²	AIC	RMSE	Adj.R ²	AIC	RMSE	Adj.R ²	AIC	RMSE
Week1	0.488	32.340	0.525	0.605	28.784	0.446	0.493	7.731	0.210	0.707	32.810	0.340
Week2	0.355	40.456	0.667	0.444	38.746	0.598	0.524	15.711	0.647	0.629	39.185	0.433
Week3	0.567	28.739	0.445	0.699	23.284	0.358	0.534	7.482	0.212	0.717	29.573	0.305
Week4	0.735	25.065	0.400	0.818	19.440	0.319	0.734	11.239	0.266	0.829	27.099	0.272
Week5	0.775	22.987	0.376	0.837	18.250	0.308	0.768	10.539	0.263	0.852	25.598	0.258
Week6	0.813	20.081	0.345	0.855	16.485	0.293	0.793	14.704	0.336	0.850	24.637	0.261
Week7	0.801	21.322	0.358	0.854	16.771	0.295	0.785	16.029	0.353	0.804	25.849	0.300
All	0.648	27.284	0.445	0.730	23.109	0.374	0.662	11.919	0.327	0.770	29.250	0.310

334

Table 8. RMSE of cross validation result (LOOCV) in Hubei province

RMSE	OLS	ESF	GWR	ESF-SVC
Week1	2.739	1.648	2.313	1.398
Week2	0.830	0.695	0.793	0.615
Week3	0.534	0.464	0.479	0.405
Week4	0.469	0.405	0.415	0.378
Week5	0.452	0.406	0.430	0.426
Week6	0.425	0.479	0.453	0.446
Week7	0.723	0.646	0.723	0.586
Avg	0.882	0.675	0.801	0.607

335

336 The LOOCV results are shown in Table 8. The ESF-SVC, GWR, ESF and OLS
 337 model's average LOOCV RMSEs were 0.607, 0.801, 0.663 and 0.882, respectively. It
 338 can be seen that the GWR, ESF, and OLS model's average LOOCV RMSEs remained
 339 almost at the same level and the ESF-SVC model outperformed the other 3 models in
 340 terms of LOOCV RMSE.

341 The Moran's I values for model residuals are displayed in Table 9. These numbers
 342 show that the OLS residuals were significantly spatially autocorrelated during the first
 343 3 weeks, but that they did not show significant spatial autocorrelation in the final weeks.
 344 The GWR residuals only showed significant spatial autocorrelation in week 2 and
 345 none of the residuals for the ESF and ESF-SVC models were spatially autocorrelated,
 346 suggesting that they could better filter out the influence of spatial autocorrelation.

To intuitively analyze model fitness, the model average residuals are visualized in Figure 3. Map legend was grade by quantiles of all 4 models' residuals. The average residuals for the ESF-SVC and GWR models remained at the same level, and were better than those of the OLS and ESF models.

Table 9. Moran's I values for model residuals in Hubei province

Moran's I	OLS	ESF	GWR	ESF-SVC
Week1	0.164*	-0.006	-0.271	-0.264
Week2	0.145*	-0.070	0.035*	-0.263
Week3	0.194**	0.011	-0.087	-0.113
Week4	0.054	-0.143	-0.034	-0.221
Week5	0.018	-0.146	-0.106	-0.266
Week6	-0.015	-0.193	-0.043	-0.245
Week7	-0.038	0.008	-0.048	-0.069

*significant at 0.1 level; **significant at 0.05 level; *** significant at 0.01 level

Figure 3 Residual maps for model fitted in Hubei province

3.1.4 ESF-SVC model coefficients

The constructed ESF-SVC model coefficients are shown in Table 10. All 10

independent variables were normalized and standardized so that the coefficient parameter values reflect how risk factors affect COVID-19 infection rates in Hubei province.

During the first three weeks, none of the 10 risk factors passed the Pearson correlation test and significance test, therefore no ESF-SVC coefficients for driving factors were obtained. PDEN and WDSP passed the modeling significance test from week 4 to week 7, with average coefficient values of 0.944 and 1.172, respectively.

In general, these results show that none of the potential risk factors but the spatial varying intercepts that generated by eigenvectors, which represent spatial autocorrelation pattern, were associated with COVID-19 infection rates in Hubei province during the first three weeks. Starting in week 4, PDEN and WDSP began to show significant impacts, but their influence did not vary spatially across 17 cities in Hubei province. WDSP, with an average coefficient of 1.172, played a more important role in the growth of the COVID-19 infection rate in Hubei province since week 4. No spatial varying coefficient maps were produced, because the health risk factor coefficients were constant.

Table 10. ESF-SVC model coefficients in Hubei province

Time		Intercept	PDEN	HOS	MS	GDP	BD	PRCP	TEMP	WDS	PRE	DEM
period										P		
Week1	coefficients	SVC	\	\	\	\	\	\	\	\	\	\
	Sig (p< 0.1)	17/17	0/17	0/17	0/17	0/17	0/17	0/17	0/17	0/362	0/17	0/17
Week2	coefficients	SVC	\	\	\	\	\	\	\	\	\	\
	Sig (p< 0.1)	17/17	0/17	0/17	0/17	0/17	0/17	0/17	0/17	0/362	0/17	0/17
Week3	coefficients	SVC	\	\	\	\	\	\	\	\	\	\
	Sig (p< 0.1)	17/17	0/17	0/17	0/17	0/17	0/17	0/17	0/17	0/17	0/17	0/17
Week4	coefficients	SVC	0.966	\	\	\	\	\	\	0.857	\	\
	Sig (p< 0.1)	17/17	17/17	0/17	0/17	0/17	0/17	0/17	17/17	17/17	0/17	0/17
Week5	coefficients	SVC	0.857	\	\	\	\	\	\	0.892	\	\
	Sig (p< 0.1)	17/17	17/17	0/17	0/17	0/17	0/17	0/17	17/17	17/17	0/17	0/17
Week6	coefficients	SVC	1.077	\	\	\	\	\	\	1.356	\	\

	Sig (p< 0.1)	17/17	17/17	0/17	0/17	0/17	0/17	0/17	17/17	17/17	0/17	0/17
Week7	coefficients	SVC	0.875	\	\	\	\	\	\	1.582	\	\
	Sig (p< 0.1)	17/17	17/17	0/17	0/17	0/17	0/17	0/17	17/17	17/17	0/17	0/17
Total	coefficients	SVC	0.944	\	\	\	\	\	\	1.172	\	\
Average	Sig (p< 0.1)	17/17	10/17	0/17	0/17	0/17	0/17	0/17	17/17	10/17	0/17	0/17

* The “\” represent for factors that did not pass significant test and not added for modeling; Sig (at 0.1): number/17 represent the number of cities whose corresponding coefficient passed the significant test at 0.1 level (p value<0.1)

3.2 Mainland China

3.2.1 Descriptive statistics

Table 11 shows the descriptive statistics for average health risk factors and COVID-19 infection rate during the study period. All variables were standardized and normalized for better comparison in modeling. Table 12 was the Moran’s I value for COVID-19 infection rate through 7 weeks from January 22 to March 10, 2020. The Moran’s I value range from 0.401 to 0.544 at a significant level of 0.001, indicating that the COVID-19 infection rate within mainland China was highly spatially autocorrelated.

385

386 Table 11. Summary of the dataset in mainland China

Value	COVID-19	PDEN	HOS	MS	GDP	BD	PRCP	TEMP	WDSP	PRE	DEM
	(case/10000 person until March 10)	(person/ km ²)	(count / km ²)			(count/ km ²)	(inch)	(°C)	(knots)	(hPa)	(m)
Min	0.000	0.29	0.00	0.00	20.95	0.001	0.00	-15.56	2.00	14.02	1.27
Max	433.309	6436.05	92.40	12.68	25123.45	2.24	44.64	16.11	191.00	1196.81	5046.47
Mean	3.860	314.56	6.01	0.05	2044.53	0.08	2.34	4.27	7.81	534.40	777.46
Std.dev	25.037	508.72	8.73	0.67	3045.12	0.18	4.68	11.92	17.81	287.98	1007.18

387

388 Table 12. Moran’s I for COVID-19 infection rate in Mainland China

Time period	Moran’s I	P’value
Week1	0.401	<0.001
Week2	0.570	<0.001
Week3	0.513	<0.001

Week4	0.542	<0.001
Week5	0.543	<0.001
Week6	0.544	<0.001
Week7	0.544	<0.001

389

390 3.2.2 Correlation analysis and multicollinearity diagnosis

391 Table 13 shows the covariates that were correlated with weekly average COVID-19
392 infection rates (WAVG). All but two of the risk factors (PRCP and WDSP) passed the
393 0.05 significance test. In general, there were positive correlations between PDEN, HOS,
394 MS, GDP, BD, and WAVG, whereas TEMP, PRE, and DEM were negatively correlated
395 with WAVG. However, the value of Pearson correlation coefficients could not fully
396 reflect the relationship between health risk factors and COVID-19 infection rates: some
397 health factors might contribute negatively in the regression models but provide positive
398 correlation coefficients when there were confounding effects (McDonald, 2009). Table
399 14 shows the VIF of risk factors. All the VIF values are less than 10, which means the
400 multicollinearity among these factors was weak and could be added for modeling.

401 Table 13. Pearson correlation coefficients between WAVG and risk factors in mainland China

Time period	PDEN	HOS	MS	GDP	BD	PRCP	TEMP	WDSP	PRE	DEM
Week1	0.249***	0.216***	0.237***	0.278***	0.234***	-0.126*	0.092*	-0.130*	0.045	-0.422**
Week2	0.231***	0.161***	0.260***	0.223***	0.209***	-0.008	0.304***	-0.044	-0.145**	-0.455**
Week3	0.236***	0.182***	0.193***	0.232***	0.207***	0.130**	0.325***	-0.035	-0.279***	-0.507**
Week4	0.234***	0.178***	0.226***	0.229***	0.209***	0.148**	0.302***	-0.082	-0.293***	-0.502**
Week5	0.234***	0.178***	0.231***	0.229***	0.209***	0.122**	0.273***	-0.039	-0.309***	-0.501**
Week6	0.234***	0.178***	0.233***	0.229***	0.208***	0.116*	0.299***	-0.042	-0.309***	-0.500**
Week7	0.233***	0.178***	0.234***	0.229***	0.208***	0.078	0.261***	-0.096	-0.317***	-0.499**

402 * significant at 0.1 level; ** significant at 0.05 level; *** significant at 0.01 level

403 Table 14. Variance Inflation Factor (VIF) of exploratory variables in mainland China

Time period	PDEN	HOS	MS	GDP	BD	PRCP	TEMP	WDSP	PRE	DEM
Week1	6.403	4.712	1.050	6.381	4.961	1.067	1.150	1.021	1.107	1.245
Week2	6.555	4.736	1.045	6.356	5.082	1.092	1.134	1.030	1.196	1.251

Week3	6.842	4.787	1.047	6.389	5.223	1.148	1.278	1.030	1.487	1.571
Week4	6.769	4.793	1.050	6.361	5.180	1.123	1.245	1.039	1.492	1.547
Week5	6.980	4.795	1.044	6.334	5.387	1.242	1.289	1.032	1.499	1.545
Week6	7.137	4.788	1.045	6.306	5.496	1.218	1.295	1.034	1.508	1.555
Week7	6.772	4.787	1.050	6.324	5.154	1.120	1.236	1.028	1.504	1.535

404

405 3.2.3 Model assessment

406 .

407 Table 15 shows the results for the 4 models. In terms of model fitness, the ESF-SVC
408 model performed the best, with an average adjusted R^2 of 0.624, which is also higher
409 than that of GWR(0.566), ESF (0.522) and OLS model (0.303). In weeks 5 and 6, the
410 ESF-SVC model's RMSE is smaller than that of the other three models. However, in
411 other weeks, the RMSE of the GWR model is the lowest. The average RMSE of the
412 OLS, ESF, GWR and ESF-SVC models are 1.601, 1.288, 1.063 and 1.134, respectively.
413 In terms of AIC, the results produced with the ESF-SVC model are not always better
414 than those produced with the other models, since the spatially varying coefficients
415 increase model complexity.

416 The LOOCV results are shown in Table 16. The ESF-SVC, GWR, ESF and OLS
417 model's average LOOCV RMSEs were 3.067, 5.732, 3.422 and 5.805, respectively.
418 Hence, the ESF-SVC model once again outperformed the other models in terms of
419 LOOCV.

420 Table 15. Model performance comparison in mainland China

	OLS			ESF			GWR			ESF-SVC		
	Adj. R^2	AIC	RMSE	Adj. R^2	AIC	RMSE	Adj. R^2	AIC	RMSE	Adj. R^2	AIC	RMSE
Week1	0.269	1208.950	1.261	0.380	1155.900	1.149	0.425	1020.251	0.871	0.480	1170.977	1.035
Week2	0.323	1266.708	1.365	0.582	1110.828	1.042	0.625	1117.044	0.816	0.667	1162.689	0.907
Week3	0.306	1486.751	1.850	0.522	1367.015	1.500	0.616	1220.195	1.162	0.622	1411.172	1.330
Week4	0.310	1423.892	1.696	0.547	1288.458	1.339	0.623	1150.409	1.050	0.653	1331.335	1.173

Week5	0.305	1418.737	1.684	0.538	1287.841	1.337	0.519	1254.774	1.280	0.651	1322.890	1.163
Week6	0.308	1414.939	1.675	0.541	1283.402	1.329	0.522	1250.099	1.272	0.647	1322.643	1.166
Week7	0.304	1414.346	1.674	0.543	1279.484	1.322	0.626	1123.939	0.993	0.646	1320.905	1.163
All	0.303	1376.332	1.601	0.522	1253.276	1.288	0.566	1162.387	1.063	0.624	1291.802	1.134

*week1: from 0122 to 0128; week2: from 0129 to 0204; ...; week7: from 0304 to 0310

Table 16. RMSEs of cross validation (LOOCV) results in mainland China

RMSE	OLS	ESF	GWR	ESF-SVC
Week1	3.583	2.689	3.788	2.560
Week2	5.801	3.407	5.327	2.997
Week3	6.476	3.678	7.412	3.267
Week4	6.273	3.581	6.243	3.184
Week5	6.181	3.579	5.500	3.164
Week6	6.154	3.532	5.653	3.139
Week7	6.167	3.486	6.200	3.160
Avg	5.805	3.422	5.732	3.067

The Moran's I values for the model residuals are displayed in Table 17. All of the OLS model residuals were significantly spatially autocorrelated, and some of the ESF (weeks 1) and GWR (weeks 5 and 6) model residuals showed significant spatial autocorrelation as well. None of the Moran's I values for the ESF-SVC model residuals passed the significance test, suggesting that the ESF-SVC model could better filter out the influence of spatial autocorrelation.

Table 17. Moran's I value for model residuals in Mainland China

Moran's I	OLS	ESF	GWR	ESF-SVC
Week1	0.194***	0.045*	-0.030	-0.027
Week2	0.290***	-0.003	-0.028	-0.070
Week3	0.246***	-0.001	-0.052	-0.077
Week4	0.279***	0.002	-0.051	-0.082
Week5	0.285***	0.012	0.064**	-0.079
Week6	0.284***	0.018	0.062**	-0.074
Week7	0.286***	0.012	-0.066	-0.070

*significant at 0.1 level; **significant at 0.05 level; *** significant at 0.01 level

To intuitively analyze the model performance, Figure 4 shows the average residuals in mainland China, which were calculated by taking the absolute average predicted

values minus the average original transformed COVID-19 infection rates, the corresponding map legend was grade by quantiles of all 4 models' residuals. The OLS model performed the worst (Figure 4a) and showed large modeling errors in Hubei province and in the west of China. The ESF model performed better and also produced large modeling errors in the west of China (Figure 4b). The mean modeling errors for the GWR and ESF-SVC models were similar across mainland China as a whole. However, the ESF-SVC model (Figure 4d) produced relatively large errors in the northwest (where there is few cities and few reported infections) while the GWR model (Figure 4c) produced relatively large modeling errors in central China, where cities with relatively higher infection rates are located.

Figure 4 Residual maps for models in mainland China

3.2.4 ESF-SVC model coefficients

The constructed ESF-SVC model coefficients are shown in Table 18. The coefficient parameter values reflect how the risk factors affected the COVID-19 infection rates throughout mainland China since all 10 of the independent variables were normalized and standardized.

MS was positively correlated with the COVID-19 infection rate, with an average spatial varying coefficient of 171.81. The weekly average MS coefficient increased at first and reached its peak of 188.48 at week 2, then it decreased to 170.41 and flattered around 171 after week 5. The number of cities whose MS coefficients passed the significance test reached the peak of 219 at week 2 and then decreased over time as well. BD was also positively and significantly correlated with the COVID-19 infection rate, with all cities passed the significant test after week 3 and an average spatial varying coefficient of 4.69. DEM was negatively correlated with the COVID-19 infection rate, with an average spatially varying coefficient of -4.02. The number of cities whose DEM coefficients passed the significance test decreased over time. Although the average spatially varying coefficient of TEMP was just -0.60, its effect on the COVID-19 infection rate was more complicated because the coefficients included a mix of positive and negative values. The number of cities whose TEMP coefficients passed the significance test was also complicated and flattered though time, ranging from 113 (week1) to 193 (week 5). PDEN and PRE only passed models' significant test in the first week and both of their coefficients were constants.

Figure 5 shows the spatially varying coefficient map of 4 selected risk factors (MS, BD, TEMP, DEM) over the 7 weeks. Figure 5 MS (a)-(g), for example, shows that MS contributed more to COVID-19 in the central and west-north cities than other parts of

mainland China. The cities with non-significant MS coefficients mostly occur near the Chinese border. BD was also positively correlated with the COVID-19 infection rate (Figure 5 BD (a)-(g)). The coefficients were significant in the whole study area and relatively larger in cities in southern China than they were in other areas after week 2. The coefficients for temperature (Figure 5 TEMP (a)-(g)) were mostly positive during the first two weeks, with an average value of 0.555. However, the number of cities with negative coefficients increased, especially in southeast and northeast China as time passed. The DEM coefficients (Figure 5 DEM (a)-(g)) were mainly negative and cities that passed the significance test and had high model coefficients after week 3 mostly occurred in Yunan, Guangxi, Guizhou, Sichuan, Neimeng, Xinjiang, Liaoning and Gansu provinces.

Table 18. ESF-SVC model coefficients in mainland China

Time period		Intercept	PDEN	HOS	MS	GDP	BD	PRCP	TEMP	WDS	PRE	DEM
										P		
Week1	coefficients	-3.37	3.86	\	SVC/148.39	\	SVC/5.76	\	SVC/0.58	\	0.070	-2.62
	Sig (p< 0.1)	362/362	14/362	0/362	211/362	0/362	265/362	0/362	112/362	0/362	11/362	362/362
Week2	coefficients	SVC	\	\	SVC/188.48	\	SVC/4.88	\	SVC/0.53	\	\	SVC/-3.89
	Sig (p< 0.1)	362/362	0/362	0/362	219/362	0/362	347/362	0/362	150/362	0/362	0/362	241/362
Week3	coefficients	SVC	\	\	SVC/177.35	\	SVC/4.76	\	SVC/-0.82	\	\	SVC/-5.04
	Sig (p< 0.1)	362/362	0/362	0/362	217/362	0/362	362/362	0/362	126/362	0/362	0/362	107/362
Week4	coefficients	SVC	\	\	SVC/176.16	\	SVC/5.06	\	SVC/-0.31	\	\	SVC/-4.32
	Sig (p< 0.1)	362/362	0/362	0/362	208/362	0/362	362/362	0/362	125/362	0/362	0/362	94/362
Week5	coefficients	SVC	\	\	SVC/170.42	\	SVC/4.83	\	SVC/-0.31	\	\	SVC/-4.11
	Sig (p< 0.1)	362/362	0/362	0/362	207/362	0/362	362/362	0/362	136/362	0/362	0/362	89/362
Week6	coefficients	SVC	\	\	SVC/171.13	\	SVC/4.79	\	SVC/-0.13	\	\	SVC/-3.96
	Sig (p< 0.1)	362/362	0/362	0/362	201/362	0/362	362/362	0/362	193/362	0/362	0/362	83/362
Week7	coefficients	SVC	\	\	SVC/170.72	\	SVC/4.59	\	SVC/-0.14	\	\	SVC/-4.22
	Sig (p< 0.1)	362/362	0/362	0/362	201/362	0/362	362/362	0/362	141/362	0/362	0/362	97/362
Total	coefficients	SVC	\	\	SVC/171.81	\	SVC/4.96	\	SVC/-0.60	\	\	SVC/-4.02
Average	Sig (p< 0.1)	362/362	2/362	0/362	208/362	0/362	346/362	0/362	138/362	0/362	2/362	153/362

* The “\” represent for factors that did not pass significant test and not added for modeling; the “svc/number” represents the average spatially varying coefficients; Sig (at 0.1): number/326 the

represent the number of cities whose corresponding coefficient passed the significant test at 0.1 level
(p value<0.1).

Figure 5 ESF-SVC spatially varying coefficient maps of health risk factors in mainland China
*each column represents a risk factor (MS, BD, TEMP, DEM) and each line (a~g) represents a week

4. Discussion

4.1 Improving model accuracy

According to the model assessment results, the newly developed ESF-SVC model performed better than the other three models in modeling COVID-19 infection rates.

When the study area was limited to Hubei province, the average adjusted R^2 of the ESF-SVC model was 0.770, which was 16.31% higher than that of GWR model (0.662), 5.48% higher than that of ESF model (0.730), and 18.83% than that of OLS model (0.648). The AIC values generated with the ESF-SVC model are larger compared to those produced with the other models, because the spatially varying coefficients increase model complexity. The cross validation results produced with the GWR, ESF, and OLS model were similar (notwithstanding minor variations) and outperformed by the ESF-SVC model. None of model residuals displayed significant spatial autocorrelation from week 4 onwards, which might be caused by the small study area and/or small sample size.

When the study area was expanded to mainland China, the average adjusted R^2 of ESF-SVC model was 0.624, which was 10.25% higher than GWR (0.566), 19.54% higher than ESF (0.522) and 105.94% higher than the OLS model (0.303). The average RMSE value of the ESF-SVC model was much higher than in the ESF and OLS models, and slightly higher than that of the GWR model. The AIC of the ESF-SVC model was no better than that of the other models, because the spatial varying coefficients increased the model complexity and reduced the degrees of freedom. The LOOCV results for the ESF-SVC model, with an average RMSE of 3.160, were significantly smaller than those for the GWR (6.200) and OLS models (6.167) and slightly smaller than that for the ESF model (3.486). This suggests that the ESF-SVC model can better estimate the relationship between COVID-19 infection rates and health risk factors in different areas. The GWR model is less robust and is plagued by multicollinearity and over-fitting problems (Helbich and Griffith, 2016; Murakami et al., 2017; Wheeler and Tiefelsdorf, 2005). The average residual maps showed that the GWR and ESF-SVC models generated similar errors in mainland China. The ESF-SVC model generated

relatively large modeling errors in northwest China, whereas the GWR model produced large errors in central China. Both models encountered problems in areas with few infections or no data for the health risk factors. The ESF-SVC model outperformed the other 3 models when only the infected areas were modeled. In addition, none of the Moran's I values generated with the ESF-SVC model residuals were significant, indicating that the ESF-SVC model could better filter out the influence of spatial autocorrelation across large areas.

Overall, the ESF-SVC method provide a more accurate path for modeling COVID-19 infection rates by simultaneously considering the spatial autocorrelation and the spatial heterogeneity. The ESF-SVC method also took good control of the over-fitting problems that plagued the GWR model, thereby providing a more robust model.

4.2 Influence of health risk factors

The spatial varying coefficient values reflect how the corresponding risk factor possibly affected the COVID-19 infection rates because the independent variables were all standardized and normalized. If the risk factor passed the significance test during the modeling, the larger the absolute coefficient, the greater its contribution to the COVID-19 infection rate.

When the study area was limited to Hubei province, only PDEN and WDSP passed the significance test and were selected for modeling. WDSP, with a larger average coefficient than that of PDEN, contributes more to the increase of infection rate after February 12. This might be because the virus could spread more in windy weather (Coşkun et al., 2021). However, as the study area expanded to mainland China, WDSP did not pass the model significance test, indicating that the influence of WDSP is more likely to act on small but high risk areas.

542 In mainland China, 8 of the 10 risk factors included in this study passed the Pearson
543 correlation test (PDEN, HOS, MS, GDP, BD, PRCP, TEMP, WDSP, PRES, DEM), and
544 6 passed the model significance test (PDEN, MS, BD, TEMP, PRES, DEM) and were
545 selected and used to develop the final models. MS had the largest influence on COVID-
546 19 infection rate, with an average coefficient of 171.81, way larger than that of building
547 density (4.69), DEM (-4.02) and temperature (-0.60). It was found that Migration Score
548 (MS) was less contributive and less significant in cities near Wuhan than cities farther
549 away. The average MS coefficients in provinces that contains or near Wuhan (Hubei,
550 Anhui, Jiangxi, Jiangsu, Zhejiang and Hunan) range from 77.75 to 132.17, with a mean
551 value of 106.16. While the average MS coefficients in provinces that are far away from
552 Wuhan (Gansu, Heilongjiang, Qinghai, Yunnan, Jilin and Xinjiang) range from 204.11
553 to 227.34, with a mean value of 214.02, which was 101.06% larger than that was in
554 provinces contains or near Wuhan. This might be because there was intense interaction
555 between Wuhan and nearby cites, be it by private car or other travel method that are not
556 recorded in the Baidu qianxi platform. When traveling to cities that are far away, people
557 are more likely to choose public transportation (airport, railway or coach), so the
558 migration score is associated with infection rates. Also, the influence of MS on COVID-
559 19 infection rate reached a peak of 188.48 at week 2 (around January 29, 2020) and
560 then decreased. This may be explained by the city lockdown policy announced at
561 January 23, 2020, around 3 weeks after the first reported case of COVID-19 in Wuhan.
562 This policy prevented people from leaving or entering Wuhan by any means of
563 transportation until April 8, 2020, before which there are no increased confirmed cases
564 within for 21 days. Public transportation, businesses as well as entertaining places
565 within Wuhan were closed to ensure rigorous home quarantine. After the announcement
566 of city lockdown policy, other cities, especially cities around Wuhan, adjusted the

emergency response level in time and applied strict social distancing policy, suggesting people to stay home as much as possible and cancel gathering and events, which largely weakened the virus spread (Bashir et al., 2020; Lau et al., 2020; W. Shi et al., 2020; Sun et al., 2020). The building density (BD) had relatively greater impact on cities in southern China since week 3 (around February 5, 2020), but the variation of coefficients between cities was small, with average value ranging from 4.41 to 4.91. This indicates that the clustering of entertainment places (restaurants, shopping malls, visiting tourists, etc.) accelerated the spread of COVID-19, but the size of this effect and its differences among cities shrank if social distancing was strong. In terms of temperature, the coefficients were mostly positive in the first 2 weeks, with an average temperature of 2.42°C. As time passed, more cities, especially those in northeast and southwest of China (e.g. cities within Guangdong, Guangxi, Yunnan and Fujian province), had negative coefficients that passed the significant test, with an average temperature of 8.17°C, suggesting that when temperature reached a certain point, the increase of temperature might result in a decrease of COVID-19 infection rate. As for altitude (DEM), cities that passed the significance test and had high model coefficients were mainly found in plateau regions with an average altitude of 1299.7m, suggesting that people living in high altitude areas who had a better tolerance to hypoxia might be more resistant to COVID-19 virus. Unlike in the extent of Hubei province, population density only showed weakly significant influence on COVID-19 infection rate at the first week and its coefficients were constants, indicating that in large study extents, population density could not explain the increase of COVID-19 infection rate well when compared with migration outflow and the clustering of buildings.

4.3 Limitations and future studies

Although the constructed ESF-SVC model performed well with improved model fitness and model robustness by considering the problem of spatial heterogeneity in the study of COVID-19, it still has several limitations. First, some risk factors that influence the spread of COVID-19, such as age structure, education level, government policy and community response, were not used. Second, although we considered the time lag effects by using the average of variables in the previous weeks, other temporal characteristics that may influence the COVID-19 infection rates, such as different temporal lag effects at day level and work with the date of the first symptoms should be considered as well. Third, the impact of driving factors may vary within different regions and in different study scale so more studies should be conducted. Another issue that should be considered is the model efficiency. Leave One Out cross-validation method was applied in our study to avoid the spatial impact on the cross-validation testing set, which would be time consuming for large dataset and needs to be improved in the future.

5. Conclusions

In this paper, we developed an ESF-SVC method for modeling COVID-19 infection rates. Population density, hospital capacity, migration score, GDP, building density, precipitation, temperature, air pressure and DEM were considered as health risk factors and the experiments was conducted using two study extents: Hubei province and mainland China over 7 weeks from January 22, 2020 to March 10, 2020.

The proposed ESF-SVC model could take good control of over-fitting problems, with a higher average adjusted R^2 compared with ESF and GWR. Also, the ESF-SVC

model's cross validation RMSE were also largely lower than other three models, indicating that it can better estimate the relationship between COVID-19 infection rate and health risk factors within different areas.

The ESF-SVC's spatial varying coefficients at different periods were displayed in tables and maps for spatial and temporal pattern discover. The effect of health risk factors was different as the study extents and study period change. In Hubei province, only wind speed (WDSP) and population density (PDEN) passed the model significant test. WDSP contributes more to the increase of infection rate than other health risk factors after February 12, indicating that the virus could spread by wind in high-risk areas with larger population density. Therefore, wearing a mask while going out is the preferred response, which was also recommended by other related studies within different study region(Biala and Khaliq, 2021; Dbouk and Drikakis, 2020; Eikenberry et al., 2020). When the study was extended to mainland China, 6 health risk factors passed the model significance test and were implied for modeling. Migration Score contributes the most to the COVID-19 infection rate followed by building density, altitude and temperature. Population density and air pressure had relatively weaker influence when compared with the other four factors. The effect of these selected health risk factors was spatially varying. Migration score was less contributive and significant in cities near Wuhan than cities farther away, while cities with larger building density coefficients were mainly found in southern China. After around January 29th, the influence of MS and BD decreased, suggesting that the rigorous lockdown and social distancing policies worked. DEM contributes to the decline of COVID-19 infection rate and its influence became more significant in high altitude cities with time passed. The influence of temperature was firstly positive correlated with COVID-19 infection rate, while after February 11th, the increase of TEMP showed weak but significant impact

on the decrease of COVID-19 infection rate in southeast and northwest cities.

In general, our method could better control multicollinearity and over-fitting problems that plague the GWR model. In addition, it provides an innovative way of exploring how health risk factors influence people differently across space and time. The findings about the impact of health risk factors were also partially consistent with previous studies on COVID-19 and other respiratory infectious diseases, which could help people and government awareness of the potential health risks and shed some light in COVID-19 control strategies. As the proposed method was not limited by datasets, it could be a reference to other spatial epidemiology researches. As a next step, we plan to add temporal variables in the spatial weights matrix and expand the ESF-SVC model into an ESF based spatiotemporal varying coefficient model. We will also consider how to expand the study area to the entire world and explore the spatiotemporal patterns within different countries and continents.

Abbreviations

OLS: Ordinary least square linear regression; ESF: Eigenvector spatial filtering regression; GWR: Geographical weighted regression; ESF-SVC: Eigenvector Spatial Filtering based spatially varying coefficient model; MS: Migration Score; PDEN: Population density; GDP: Gross domestic; HOS: Hospital capacity; BD: Building Density; TEMP: average temperature; PRCP: precipitation; WDSP: wind speed; PRE: air pressure; DEM: altitude; AIC: Akaike information criterion; RMSE: Root Mean Squared Error

Acknowledgments

This work was supported by the National Key S&T Special Projects of China [Grant No. 2017YFB0503704]. The authors sincerely acknowledge their financial support

for this research.

Authors' contributions

MC, YC and J.P.W conceptualized the study; MC and YC designed the methodology; MC, HT and TC collected and curated the data; MC and TC wrote and maintained the code; MC wrote the first version of the manuscript while YC and J.P.W edited the final manuscript; All authors read and approved the final manuscript.

Funding

This work was supported by the National Key S&T Special Projects of China [Grant No. 2017YFB0503704]. The authors sincerely acknowledge their financial support for this research.

Availability of data and materials

All data and materials used are public accessed as illustrated in the article.

Declarations

Ethics approval and consent to participate:

Not applicable

Consent for publication:

Not applicable

Competing interests:

The authors declare no competing interest

References

- Alzahrani, A.K., Razzaq, O.A., Khan, N.A., Alshomrani, A.S., Ullah, M.Z., 2021. Transmissibility of epidemic diseases caused by delay with local proportional fractional derivative. *Adv. Differ. Equations* 2021, 1–20.
- Andersen, L.M., Harden, S.R., Sugg, M.M., Runkle, J.D., Lundquist, T.E., 2021.

689 Analyzing the spatial determinants of local Covid-19 transmission in the United
690 States. *Sci. Total Environ.* 754, 142396.
691 <https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.142396>

692 ArunKumar, K.E., Kalaga, D. V, Sai Kumar, C.M., Chilkoor, G., Kawaji, M., Brenza,
693 T.M., 2021. Forecasting the dynamics of cumulative COVID-19 cases
694 (confirmed, recovered and deaths) for top-16 countries using statistical machine
695 learning models: Auto-Regressive Integrated Moving Average (ARIMA) and
696 Seasonal Auto-Regressive Integrated Moving Averag. *Appl. Soft Comput.* 103,
697 107161. <https://doi.org/https://doi.org/10.1016/j.asoc.2021.107161>

698 Bashir, M.F., MA, B., Shahzad, L., 2020. A brief review of socio-economic and
699 environmental impact of Covid-19. *Air Qual. Atmos. Heal.* 13, 1403–1409.
700 <https://doi.org/10.1007/s11869-020-00894-8>

701 Beale, L., Abellan, J., Hodgson, S., Jarup, L., 2008. Methodologic Issues and
702 Approaches to Spatial Epidemiology. *Environ. Health Perspect.* 116, 1105–1110.
703 <https://doi.org/10.1289/ehp.10816>

704 Biala, T.A., Khaliq, A.Q.M., 2021. A fractional-order compartmental model for the
705 spread of the COVID-19 pandemic. *Commun. Nonlinear Sci. Numer. Simul.* 98,
706 105764. <https://doi.org/https://doi.org/10.1016/j.cnsns.2021.105764>

707 Box, G.E.P., Cox, D.R., 1964. An Analysis of Transformations. *J. R. Stat. Soc. Ser. B*
708 26, 211–243. <https://doi.org/https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>

709 Brenning, A., 2012. Spatial cross-validation and bootstrap for the assessment of
710 prediction rules in remote sensing: The R package sperrorest, in: 2012 IEEE
711 International Geoscience and Remote Sensing Symposium. pp. 5372–5375.
712 <https://doi.org/10.1109/IGARSS.2012.6352393>

713 Brunsdon, C., Fotheringham, A.S., Charlton, M.E., 1996. Geographically Weighted

714 Regression: A Method for Exploring Spatial Nonstationarity. *Geogr. Anal.* 28,
715 281–298. <https://doi.org/https://doi.org/10.1111/j.1538-4632.1996.tb00936.x>

716 Ceylan, Z., 2020. Estimation of COVID-19 prevalence in Italy, Spain, and France.
717 *Sci. Total Environ.* 729, 138817.
718 <https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.138817>

719 Chadsuthi, S., Modchang, C., 2021. Modelling the effectiveness of intervention
720 strategies to control COVID-19 outbreaks and estimating healthcare demand in
721 Germany. *Public Heal. Pract.* 2, 100121.
722 <https://doi.org/https://doi.org/10.1016/j.puhip.2021.100121>

723 Chang, S., Pierson, E., Koh, P.W., Gerardin, J., Redbird, B., Grusky, D., Leskovec, J.,
724 2021. Mobility network models of COVID-19 explain inequities and inform
725 reopening. *Nature* 589, 82–87. <https://doi.org/10.1038/s41586-020-2923-3>

726 Chen, Y., Wang, B., Liu, X., Li, X., 2018. Mapping the spatial disparities in urban
727 health care services using taxi trajectories data. *Trans. GIS* 22, 602–615.
728 <https://doi.org/https://doi.org/10.1111/tgis.12333>

729 Coşkun, H., Yıldırım, N., Gündüz, S., 2021. The spread of COVID-19 virus through
730 population density and wind in Turkey cities. *Sci. Total Environ.* 751, 141663.
731 <https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.141663>

732 Cucinotta, D., Vanelli, M., 2020. WHO Declares COVID-19 a Pandemic. *Acta*
733 *Biomed.* 91, 157–160. <https://doi.org/10.23750/abm.v91i1.9397>

734 Damialis, A., Gilles, S., Sofiev, M., Sofieva, V., Kolek, F., Bayr, D., Plaza, M.P.,
735 Leier-Wirtz, V., Kaschuba, S., Ziska, L.H., Bielory, L., Makra, L., del Mar
736 Trigo, M., Traidl-Hoffmann, C., 2021. Higher airborne pollen concentrations
737 correlated with increased SARS-CoV-2 infection rates, as evidenced from 31
738 countries across the globe. *Proc. Natl. Acad. Sci.* 118, e2019034118.

739 <https://doi.org/10.1073/pnas.2019034118>

740 Dbouk, T., Drikakis, D., 2020. Weather impact on airborne coronavirus survival.

741 *Phys. Fluids* 32, 93312. <https://doi.org/10.1063/5.0024272>

742 Dellicour, S., Linard, C., Van Goethem, N., Da Re, D., Artois, J., Bihin, J., Schaus, P.,

743 Massonnet, F., Van Oyen, H., Vanwambeke, S.O., Speybroeck, N., Gilbert, M.,

744 2021. Investigating the drivers of the spatio-temporal heterogeneity in COVID-

745 19 hospital incidence—Belgium as a study case. *Int. J. Health Geogr.* 20, 29.

746 <https://doi.org/10.1186/s12942-021-00281-1>

747 Desai, P.S., 2021. News Sentiment Informed Time-series Analyzing AI (SITALA) to

748 curb the spread of COVID-19 in Houston. *Expert Syst. Appl.* 180, 115104.

749 <https://doi.org/https://doi.org/10.1016/j.eswa.2021.115104>

750 Dowd, J.B., Andriano, L., Brazel, D.M., Rotondi, V., Block, P., Ding, X., Liu, Y.,

751 Mills, M.C., 2020. Demographic science aids in understanding the spread and

752 fatality rates of COVID-19. *Proc. Natl. Acad. Sci.* 117, 9696 LP – 9698.

753 <https://doi.org/10.1073/pnas.2004911117>

754 Eikenberry, S.E., Mancuso, M., Iboi, E., Phan, T., Eikenberry, K., Kuang, Y.,

755 Kostelich, E., Gumel, A.B., 2020. To mask or not to mask: Modeling the

756 potential for face mask use by the general public to curtail the COVID-19

757 pandemic. *Infect. Dis. Model.* 5, 293–308.

758 <https://doi.org/https://doi.org/10.1016/j.idm.2020.04.001>

759 EL Azzaoui, A., Singh, S.K., Park, J.H., 2021. SNS Big Data Analysis Framework for

760 COVID-19 Outbreak Prediction in Smart Healthy City. *Sustain. Cities Soc.* 71,

761 102993. <https://doi.org/https://doi.org/10.1016/j.scs.2021.102993>

762 Elliott, P., Wartenberg, D., 2004. Spatial epidemiology: Current approaches and

763 future challenges. *Environ. Health Perspect.* 112, 998–1006.

764 <https://doi.org/10.1289/ehp.6735>

765 Fanelli, D., Piazza, F., 2020. Analysis and forecast of COVID-19 spreading in China,
 766 Italy and France. *Chaos, Solitons & Fractals* 134, 109761.
 767 <https://doi.org/https://doi.org/10.1016/j.chaos.2020.109761>

768 Fotheringham, A.S., Yang, W., Kang, W., 2017. Multiscale Geographically Weighted
 769 Regression (MGWR). *Ann. Am. Assoc. Geogr.* 107, 1247–1265.
 770 <https://doi.org/10.1080/24694452.2017.1352480>

771 Gerkman, L.M., Ahlgren, N., 2014. Practical Proposals for Specifying k-Nearest
 772 Neighbours Weights Matrices. *Spat. Econ. Anal.* 9, 260–283.
 773 <https://doi.org/10.1080/17421772.2014.930167>

774 Ghojogh, B., Crowley, M., 2019. The Theory Behind Overfitting, Cross Validation,
 775 Regularization, Bagging, and Boosting: Tutorial.

776 Griffith, D.A., 2008. Spatial-Filtering-Based Contributions to a Critique of
 777 Geographically Weighted Regression (GWR). *Environ. Plan. A Econ. Sp.* 40,
 778 2751–2769. <https://doi.org/10.1068/a38218>

779 Guleryuz, D., 2021. Forecasting outbreak of COVID-19 in Turkey; Comparison of
 780 Box–Jenkins, Brown’s exponential smoothing and long short-term memory
 781 models. *Process Saf. Environ. Prot.* 149, 927–935.
 782 <https://doi.org/https://doi.org/10.1016/j.psep.2021.03.032>

783 Gupta, S., Raghuwanshi, G.S., Chanda, A., 2020. Effect of weather on COVID-19
 784 spread in the US: A prediction model for India in 2020. *Sci. Total Environ.* 728,
 785 138860. <https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.138860>

786 Han, Y., Yang, L., Jia, K., Li, J., Feng, S., Chen, W., Zhao, W., Pereira, P., 2021.
 787 Spatial distribution characteristics of the COVID-19 pandemic in Beijing and its
 788 relationship with environmental factors. *Sci. Total Environ.* 761, 144257.

789 <https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.144257>

790 Hawkins, R.B., Charles, E.J., Mehaffey, J.H., 2020. Socio-economic status and

791 COVID-19–related cases and fatalities. *Public Health* 189, 129–134.

792 <https://doi.org/https://doi.org/10.1016/j.puhe.2020.09.016>

793 He, J., Chen, G., Jiang, Y., Jin, R., Shortridge, A., Agusti, S., He, M., Wu, J., Duarte,

794 C.M., Christakos, G., 2020. Comparative infection modeling and control of

795 COVID-19 transmission patterns in China, South Korea, Italy and Iran. *Sci.*

796 *Total Environ.* 747, 141447.

797 <https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.141447>

798 Helbich, M., Griffith, D.A., 2016. Spatially varying coefficient models in real estate:

799 Eigenvector spatial filtering and alternative approaches. *Comput. Environ. Urban*

800 *Syst.* 57, 1–11.

801 <https://doi.org/https://doi.org/10.1016/j.compenvurbsys.2015.12.002>

802 Jia, P., Dong, W., Yang, S., Zhan, Z., Tu, L., Lai, S., 2020. Spatial Lifecourse

803 Epidemiology and Infectious Disease Research. *Trends Parasitol.* 36.

804 <https://doi.org/10.1016/j.pt.2019.12.012>

805 Jia, P., Lakerveld, J., Wu, J., Stein, A., Root, E., Sabel, C., Vermeulen, R., Remais, J.,

806 Chen, X., Brownson, R., Amer, S., Xiao, Q., Wang, L., Verschuren, W., Wu, T.,

807 Wang, Y., James, P., 2019. Top 10 Research Priorities in Spatial Lifecourse

808 Epidemiology. *Environ. Health Perspect.* 127, 74501.

809 <https://doi.org/10.1289/EHP4868>

810 Karaye, I.M., Horney, J.A., 2020. The Impact of Social Vulnerability on COVID-19

811 in the U.S.: An Analysis of Spatially Varying Relationships. *Am. J. Prev. Med.*

812 59, 317–325. <https://doi.org/https://doi.org/10.1016/j.amepre.2020.06.006>

813 Khedher, N. ben, Kolsi, L., Alsaif, H., 2021. A multi-stage SEIR model to predict the

814 potential of a new COVID-19 wave in KSA after lifting all travel restrictions.
815 Alexandria Eng. J. 60, 3965–3974.
816 <https://doi.org/https://doi.org/10.1016/j.aej.2021.02.058>

817 Kim, B., Rundle, A.G., Goodwin, A.T.S., Morrison, C.N., Branas, C.C., El-Sadr, W.,
818 Duncan, D.T., 2021. COVID-19 testing, case, and death rates and spatial socio-
819 demographics in New York City: An ecological analysis as of June 2020. Health
820 Place 68, 102539.
821 <https://doi.org/https://doi.org/10.1016/j.healthplace.2021.102539>

822 Kulu, H., Dorey, P., 2021. Infection rates from Covid-19 in Great Britain by
823 geographical units: A model-based estimation from mortality data. Health Place
824 67, 102460. <https://doi.org/https://doi.org/10.1016/j.healthplace.2020.102460>

825 Kwok, C.Y.T., Wong, M.S., Chan, K.L., Kwan, M.-P., Nichol, J.E., Liu, C.H., Wong,
826 J.Y.H., Wai, A.K.C., Chan, L.W.C., Xu, Y., Li, H., Huang, J., Kan, Z., 2021.
827 Spatial analysis of the impact of urban geometry and socio-demographic
828 characteristics on COVID-19, a study in Hong Kong. Sci. Total Environ. 764,
829 144455. <https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.144455>

830 Lau, H., Khosrawipour, V., Kocbach, P., Mikolajczyk, A., Schubert, J., Bania, J.,
831 Khosrawipour, T., 2020. The positive impact of lockdown in Wuhan on
832 containing the COVID-19 outbreak in China. J. Travel Med. 27, taaa037.
833 <https://doi.org/10.1093/jtm/taaa037>

834 Lesage, J.P., Fischer, M.M., 2008. Spatial Growth Regressions: Model Specification,
835 Estimation and Interpretation. Spat. Econ. Anal. 3, 275–304.
836 <https://doi.org/10.1080/17421770802353758>

837 Li, Juan, Yuan, P., Heffernan, J., Zheng, T., Ogden, N., Sander, B., Li, Jun, Li, Q.,
838 Bélair, J., Kong, J.D., Aruffo, E., Tan, Y., Jin, Z., Yu, Y., Fan, M., Cui, J.,

839 Tengb, Z., Zhua, H., 2020. Fangcang shelter hospitals during the COVID-19
840 epidemic, Wuhan, China. *World Heal. Organ. Bull. World Heal. Organ.* 98, 830.
841 <https://doi.org/http://dx.doi.org/10.2471/BLT.20.258152>

842 Liao, Zhifang, Lan, P., Liao, Zhining, Zhang, Y., Liu, S., 2020. TW-SIR: time-
843 window based SIR for COVID-19 forecasts. *Sci. Rep.* 10, 22454.
844 <https://doi.org/10.1038/s41598-020-80007-8>

845 Lim, Y.K., Kweon, O.J., Kim, H.R., Kim, T.-H., Lee, M.-K., 2021. The impact of
846 environmental variables on the spread of COVID-19 in the Republic of Korea.
847 *Sci. Rep.* 11, 5977. <https://doi.org/10.1038/s41598-021-85493-y>

848 Liu, Z., Magal, P., Webb, G., 2020. Predicting the number of reported and unreported
849 cases for the COVID-19 epidemics in China, South Korea, Italy, France,
850 Germany and United Kingdom. <https://doi.org/10.1101/2020.04.09.20058974>

851 Ma, Y., Cheng, B., Shen, J., Wang, H., Feng, F., Zhang, Y., Jiao, H., 2021.
852 Association between environmental factors and COVID-19 in Shanghai, China.
853 *Environ. Sci. Pollut. Res.* <https://doi.org/10.1007/s11356-021-13834-5>

854 Ma, Y., Zhao, Y., Liu, J., He, X., Wang, B., Fu, S., Yan, J., Niu, J., Zhou, J., Luo, B.,
855 2020. Effects of temperature variation and humidity on the death of COVID-19
856 in Wuhan, China. *Sci. Total Environ.* 724, 138226.
857 <https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.138226>

858 Maiti, A., Zhang, Q., Sannigrahi, S., Pramanik, S., Chakraborti, S., Cerda, A., Pilla,
859 F., 2021. Exploring spatiotemporal effects of the driving factors on COVID-19
860 incidences in the contiguous United States. *Sustain. Cities Soc.* 68, 102784.
861 <https://doi.org/https://doi.org/10.1016/j.scs.2021.102784>

862 Malki, Z., Atlam, E.-S., Ewis, A., Dagnew, G., Alzighaibi, A.R., ELmarhomy, G.,
863 Elhosseini, M.A., Hassanien, A.E., Gad, I., 2020. ARIMA models for predicting

864 the end of COVID-19 pandemic and the risk of second rebound. *Neural Comput.*
865 *Appl.* <https://doi.org/10.1007/s00521-020-05434-0>

866 Mansour, S., Al Kindi, A., Al-Said, Alkhattab, Al-Said, Adham, Atkinson, P., 2021.
867 Sociodemographic determinants of COVID-19 incidence rates in Oman:
868 Geospatial modelling using multiscale geographically weighted regression
869 (MGWR). *Sustain. Cities Soc.* 65, 102627.
870 <https://doi.org/https://doi.org/10.1016/j.scs.2020.102627>

871 Martins, L.D., da Silva, I., Batista, W.V., Andrade, M. de F., Freitas, E.D. de,
872 Martins, J.A., 2020. How socio-economic and atmospheric variables impact
873 COVID-19 and influenza outbreaks in tropical and subtropical regions of Brazil.
874 *Environ. Res.* 191, 110184.
875 <https://doi.org/https://doi.org/10.1016/j.envres.2020.110184>

876 Mateo-Urdiales, A., Fabiani, M., Rosano, A., Vescio, M.F., Del Manso, M., Bella, A.,
877 Riccardo, F., Pezzotti, P., Regidor, E., Andrianou, X., 2021. Socioeconomic
878 patterns and COVID-19 outcomes before, during and after the lockdown in Italy
879 (2020). *Health Place* 71, 102642.
880 <https://doi.org/https://doi.org/10.1016/j.healthplace.2021.102642>

881 McDonald, J.H., 2009. *Handbook of Biological Statistics* 2nd Edition, Sparky House
882 Publishing Baltimore.

883 Messner, W., 2020. The Institutional and Cultural Context of Cross-National
884 Variation in COVID-19 Outbreaks. *medRxiv* 2020.03.30.20047589.
885 <https://doi.org/10.1101/2020.03.30.20047589>

886 Mollalo, A., Vahedi, B., Rivera, K.M., 2020. GIS-based spatial modeling of COVID-
887 19 incidence rate in the continental United States. *Sci. Total Environ.* 728,
888 138884. <https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.138884>

889 Murakami, D., 2017. Spatial regression modeling using the spmoran package: Boston
890 housing price data examples.

891 Murakami, D., Yoshida, T., Seya, H., Griffith, D.A., Yamagata, Y., 2017. A Moran
892 coefficient-based mixed effects approach to investigate spatially varying
893 relationships. *Spat. Stat.* 19, 68–89.
894 <https://doi.org/https://doi.org/10.1016/j.spasta.2016.12.001>

895 Niazkar, H.R., Niazkar, M., 2020. Application of artificial neural networks to predict
896 the COVID-19 outbreak. *Glob. Heal. Res. Policy* 5, 50.
897 <https://doi.org/10.1186/s41256-020-00175-y>

898 Picard, R.R., Cook, R.D., 1984. Cross-Validation of Regression Models. *J. Am. Stat.*
899 *Assoc.* 79, 575–583. <https://doi.org/10.1080/01621459.1984.10478083>

900 Prata, D.N., Rodrigues, W., Bermejo, P.H., 2020. Temperature significantly changes
901 COVID-19 transmission in (sub)tropical cities of Brazil. *Sci. Total Environ.* 729,
902 138862. <https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.138862>

903 Qi, H., Xiao, S., Shi, R., Ward, M.P., Chen, Y., Tu, W., Su, Q., Wang, W., Wang, X.,
904 Zhang, Z., 2020. COVID-19 transmission in Mainland China is associated with
905 temperature and humidity: A time-series analysis. *Sci. Total Environ.* 728,
906 138778. <https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.138778>

907 Rahman, M., Islam, M., Shimanto, M.H., Ferdous, J., Rahman, A.A.-N.S., Sagor,
908 P.S., Chowdhury, T., 2020. A global analysis on the effect of temperature, socio-
909 economic and environmental factors on the spread and mortality rate of the
910 COVID-19 pandemic. *Environ. Dev. Sustain.* [https://doi.org/10.1007/s10668-](https://doi.org/10.1007/s10668-020-01028-x)
911 [020-01028-x](https://doi.org/10.1007/s10668-020-01028-x)

912 Ravindra, K., Goyal, A., Mor, S., 2021. Does airborne pollen influence COVID-19
913 outbreak? *Sustain. Cities Soc.* 70, 102887.

914 <https://doi.org/https://doi.org/10.1016/j.scs.2021.102887>

915 Refaeilzadeh, P., Tang, L., Liu, H., 2009. Cross-Validation BT - Encyclopedia of
 916 Database Systems, in: LIU, L., ÖZSU, M.T. (Eds.), . Springer US, Boston, MA,
 917 pp. 532–538. https://doi.org/10.1007/978-0-387-39940-9_565

918 Reiner, R., Barber, R., Collins, J., Zheng, P., Adolph, C., Albright, J., Antony, C.,
 919 Aravkin, A., Bachmeier, S., Bang-Jensen, B., Bannick, M., Bloom, S., Carter,
 920 A., Castro, E., Causey, K., Chakrabarti, S., Charlson, F., Cogen, R., Combs, E.,
 921 Murray, C., 2021. Modeling COVID-19 scenarios for the United States. *Nat.*
 922 *Med.* 27. <https://doi.org/10.1038/s41591-020-1132-9>

923 Rocklöv, J., Sjödin, H., 2020. High population densities catalyse the spread of
 924 COVID-19. *J. Travel Med.* 27. <https://doi.org/10.1093/jtm/taaa038>

925 Rogerson, P., 2001. *Statistical Methods for Geography*.
 926 <https://doi.org/10.4135/9781849209953>

927 Runkle, J.D., Sugg, M.M., Leeper, R.D., Rao, Y., Matthews, J.L., Rennie, J.J., 2020.
 928 Short-term effects of specific humidity and temperature on COVID-19 morbidity
 929 in select US cities. *Sci. Total Environ.* 740, 140093.
 930 <https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.140093>

931 Saba, A.I., Elsheikh, A.H., 2020. Forecasting the prevalence of COVID-19 outbreak
 932 in Egypt using nonlinear autoregressive artificial neural networks. *Process Saf.*
 933 *Environ. Prot.* 141, 1–8.
 934 <https://doi.org/https://doi.org/10.1016/j.psep.2020.05.029>

935 Şahin, M., 2020. Impact of weather on COVID-19 pandemic in Turkey. *Sci. Total*
 936 *Environ.* 728, 138810.
 937 <https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.138810>

938 Sannigrahi, S., Pilla, F., Basu, B., Basu, A., Mölter, A., 2020. Examining the

939 association between socio-demographic composition and COVID-19 fatalities in
 940 the European region using spatial regression approach. *Sustain. Cities Soc.* 62.
 941 <https://doi.org/10.1016/j.scs.2020.102418>

942 Scarpone, C., Brinkmann, S.T., Große, T., Sonnenwald, D., Fuchs, M., Walker, B.B.,
 943 2020. A multimethod approach for county-scale geospatial analysis of emerging
 944 infectious diseases: a cross-sectional case study of COVID-19 incidence in
 945 Germany. *Int. J. Health Geogr.* 19, 32. [https://doi.org/10.1186/s12942-020-](https://doi.org/10.1186/s12942-020-00225-1)
 946 00225-1

947 Schneider, K., Skevas, I., Oude Lansink, A., 2021. Spatial Spillovers on Input-specific
 948 Inefficiency of Dutch Arable Farms. *J. Agric. Econ.* 72, 224–243.
 949 <https://doi.org/https://doi.org/10.1111/1477-9552.12400>

950 Shi, P., Dong, Y., Yan, H., Zhao, C., Li, X., Liu, W., He, M., Tang, S., Xi, S., 2020.
 951 Impact of temperature on the dynamics of the COVID-19 outbreak in China. *Sci.*
 952 *Total Environ.* 728, 138890.
 953 <https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.138890>

954 Shi, W., Tong, C., Zhang, A., Wang, B., Shi, Z., Yao, Y., Jia, P., 2020. Forecasting
 955 COVID-19 onset risk and evaluating spatiotemporal variations of the lockdown
 956 effect in China. <https://doi.org/10.21203/rs.3.rs-28675/v1>

957 Sirkeci, I.(1), Murat Yüceşahin, M.(2), 2020. Coronavirus and migration: Analysis
 958 of human mobility and the spread of covid-19. *Migr. Lett.* 17, 379–398.
 959 <https://doi.org/10.33182/ml.v17i2.935>

960 Snyder, B.F., Parks, V., 2020. Spatial variation in socio-ecological vulnerability to
 961 Covid-19 in the contiguous United States. *Health Place* 66, 102471.
 962 <https://doi.org/https://doi.org/10.1016/j.healthplace.2020.102471>

963 Sun, Z., Zhang, H., Yang, Y., Wan, H., Wang, Y., 2020. Impacts of geographic

964 factors and population density on the COVID-19 spreading under the lockdown
 965 policies of China. *Sci. Total Environ.* 746, 141347.
 966 <https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.141347>
 967 Tan, H., Chen, Y., Wilson, J.P., Zhang, J., Cao, J., Chu, T., 2020. An eigenvector
 968 spatial filtering based spatially varying coefficient model for PM2.5
 969 concentration estimation: A case study in Yangtze River Delta region of China.
 970 *Atmos. Environ.* 223, N.PAG-N.PAG.
 971 <https://doi.org/10.1016/j.atmosenv.2019.117205>
 972 Tian, S., Hu, N., Lou, J., Chen, K., Kang, X., Xiang, Z., Chen, H., Wang, D., Liu, N.,
 973 Liu, D., Chen, G., Zhang, Y., Li, D., Li, J., Lian, H., Niu, S., Zhang, L., Zhang,
 974 J., 2020. Characteristics of COVID-19 infection in Beijing. *J. Infect.* 80, 401–
 975 406. <https://doi.org/https://doi.org/10.1016/j.jinf.2020.02.018>
 976 Tobler, W.R., 1970. A Computer Movie Simulating Urban Growth in the Detroit
 977 Region. *Econ. Geogr.* 46, 234. <https://doi.org/10.2307/143141>
 978 Wheeler, D., Tiefelsdorf, M., 2005. Multicollinearity and correlation among local
 979 regression coefficients in geographically weighted regression. *J. Geogr. Syst.* 7,
 980 161–187. <https://doi.org/10.1007/s10109-005-0155-6>
 981 WHO, 2020. Weekly epidemiological update - 27 December 2020. Genève, Switz. 17.
 982 Wu, X., Yin, J., Li, C., Xiang, H., Lv, M., Guo, Z., 2020. Natural and human
 983 environment interactively drive spread pattern of COVID-19: A city-level
 984 modeling study in China. *Sci. Total Environ.* 756, 143343.
 985 <https://doi.org/10.1016/j.scitotenv.2020.143343>
 986 Yang, Z., Zeng, Z., Wang, K., Wong, S.-S., Liang, W., Zanin, M., Liu, P., Cao, X.,
 987 Gao, Z., Mai, Z., Liang, J., Liu, X., Li, S., Li, Y., Ye, F., Guan, W., Yang, Y., Li,
 988 F., Luo, S., He, J., 2020. Modified SEIR and AI prediction of the epidemics

989 trend of COVID-19 in China under public health interventions. *J. Thorac. Dis.*
 990 12, 165–174. <https://doi.org/10.21037/jtd.2020.02.64>
 991 Yin, H., Sun, T., Yao, L., Jiao, Y., Ma, L., Lin, L., Graff, J.C., Aleya, L.,
 992 Postlethwaite, A., Gu, W., Chen, H., 2021. Association between population
 993 density and infection rate suggests the importance of social distancing and travel
 994 restriction in reducing the COVID-19 pandemic. *Environ. Sci. Pollut. Res.*
 995 <https://doi.org/10.1007/s11356-021-12364-4>
 996 Zhou, K., Yang, S., Jia, P., 2020. Towards precision management of cardiovascular
 997 patients with COVID-19 to reduce mortality. *Prog. Cardiovasc. Dis.* 63, 529–
 998 530. <https://doi.org/https://doi.org/10.1016/j.pcad.2020.04.012>
 999 Zhu, Y., Xie, J., Huang, F., Cao, L., 2020. Association between short-term exposure
 1000 to air pollution and COVID-19 infection: Evidence from China. *Sci. Total*
 1001 *Environ.* 727, 138704.
 1002 <https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.138704>
 1003