



# Spatiotemporal cluster analysis of COVID-19 and its relationship with environmental factors at the city level in mainland China

Shu-qin Yang<sup>1</sup> · Zheng-gang Fang<sup>1</sup> · Cai-xia Lv<sup>1</sup> · Shu-yi An<sup>2</sup> · Peng Guan<sup>1</sup> · De-sheng Huang<sup>3</sup> · Wei Wu<sup>1</sup>

Received: 5 July 2021 / Accepted: 14 September 2021 / Published online: 30 September 2021  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

## Abstract

This study sought to identify the spatial, temporal, and spatiotemporal clusters of COVID-19 cases in 366 cities in mainland China with the highest risks and to explore the possible influencing factors of imported risks and environmental factors on the spatiotemporal aggregation, which would be useful to the design and implementation of critical preventative measures. The retrospective analysis of temporal, spatial, and spatiotemporal clustering of COVID-19 during the period (January 15 to February 25, 2020) was based on Kulldorff's time-space scanning statistics using the discrete Poisson probability model, and then the logistic regression model was used to evaluate the impact of imported risk and environmental factors on spatiotemporal aggregation. We found that the spatial distribution of COVID-19 cases was nonrandom; the Moran's  $I$  value ranged from 0.017 to 0.453 ( $P < 0.001$ ). One most likely cluster and three secondary likely clusters were discovered in spatial cluster analysis. The period from February 2 to February 9, 2020, was identified as the most likely cluster in the temporal cluster analysis. One most likely cluster and seven secondary likely clusters were discovered in spatiotemporal cluster analysis. Imported risk, humidity, and inhalable particulate matter  $PM_{2.5}$  had a significant impact on temporal and spatial accumulation, and temperature and  $PM_{10}$  had a low correlation with the spatiotemporal aggregation of COVID-19. The information is useful for health departments to develop a better prevention strategy and potentially increase the effectiveness of public health interventions.

**Keywords** COVID-19 · Cluster analysis · Geographic information system

## Introduction

The new coronavirus disease (COVID-19), which first appeared in Wuhan, is an infectious disease caused by a coronavirus called severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Lu et al. 2020), and it is widespread throughout the country and endemic in many countries and regions worldwide. COVID-19 represents another serious public health emergency since the outbreak of SARS. As of

December 22, 2020, a total of 76,250,431 cumulative cases and 1,699,230 deaths have been reported globally, including 95,916 cumulative cases and 4772 deaths in China ("<https://covid19.who.int/table>"). The disease was confirmed to be highly contagious from humans to humans, and human activity is the main cause of its spread (Coccia 2020b). In addition, short-term exposure to environmental pollutants has been recently studied to accelerate the spread of SARS-CoV-2 (Coccia 2020a; Coccia 2021a). Most infected patients have the common symptom of high fever, and some have dyspnea with chest radiograph results showing invasive lesions in both lungs (Chan et al. 2020; Huang et al. 2020). The novel coronavirus epidemic also involves a continuous process of epidemic focus, which is based on both the growth over time and the spread of space.

The occurrence of COVID-19 is regular in space, time, and space-time. The cluster patterns of COVID-19 are the focus of our research establishing the risk factors involved in the spread of COVID-19 to effectively intervene and control COVID-19. Obviously, scan statistics are excellent methods to solve multiple testing problems when two dimensions of

---

Responsible Editor: Lotfi Aleya

✉ Wei Wu  
wuwei@cmu.edu.cn

- <sup>1</sup> Department of Epidemiology, School of Public Health, China Medical University, Shenyang, Liaoning, China
- <sup>2</sup> Liaoning Provincial Center for Disease Control and Prevention, Shenyang, Liaoning, China
- <sup>3</sup> Department of Mathematics, School of Fundamental Sciences, China Medical University, Shenyang, Liaoning, China

spatial area and/or time interval are evaluated simultaneously. The scan statistics software SaTScan is mainly applied to perform temporal, spatial, and spatiotemporal analyses for many infectious diseases including schistosomiasis (Yu-Wan et al. 2019); malaria (Yan-Feng et al. 2019); dengue fever (Lai et al. 2018); hand, foot, and mouth disease (Tian et al. 2018); and hemorrhagic fever with renal syndrome (Wu et al. 2011).

In previous studies, spatiotemporal analysis methods have been widely used in the epidemiology of infectious diseases (Fang et al. 2009; Fosgate et al. 2002). Spatiotemporal analysis methods can reveal the time trends and spatiotemporal patterns of diseases from different spatial scales and can more intuitively clarify the epidemiological characteristics and epidemiological laws of infectious diseases. However, to the best of our knowledge, no specific research has been conducted on the spatiotemporal pattern of COVID-19 in mainland China. A better comprehension of the spatiotemporal distribution of COVID-19 would benefit the identification of the cities, populations, and times at highest risk, which would aid in the effective implementation of relevant preventative measures.

Many studies have shown that meteorological factors are closely related to the occurrence and spread of infectious diseases (Irfan et al. 2021; Islam et al. 2021). For example, temperature and humidity are related to the spread of severe acute respiratory syndrome (SARS)(Rahimi et al. 2021; Rosario et al. 2020; Sarkodie and Owusu 2020); some scholars believe that the spread of influenza will increase around cooler, drier air (Haque and Rahman 2020; Shen et al. 2021), and other scholars believe that the high wind speeds that cause atmospheric instability seem to reduce the number of infected people and promote the spread of air pollution (Mario Coccia 2021; Coccia 2021b). In addition, air pollution has been also widely considered to be an important factor in increasing the risk of adverse health outcomes (Du et al. 2021; Tian et al. 2021), such as ischemic heart disease, chronic obstructive pulmonary disease, and respiratory infection (Manisalidis et al. 2020; Tian et al. 2021; Wu et al. 2016). Emerging evidence suggests potential links between exposure to polluted air and the epidemic of COVID-19(Chen et al. 2021; Chen et al. 2020; Konstantinoudis et al. 2021). Research has shown that the airborne transmission is highly virulent (Srivastava 2021), particulate matter can be used as a carrier to carry viruses to spread it everywhere (Vlachogiannis et al. 2021; Yao et al. 2020), and respiratory infectious disease viruses that are efficiently transmitted through particulate matter can remain stable for a long period of time.

In this study, geographic information system (GIS)–based analyses were conducted to characterize the spatiotemporal patterns of COVID-19 in mainland China using surveillance data from January 15 to February 25, 2020, to identify spatiotemporal clusters of COVID-19 cases at the city level and to explore the influences of short-term exposure to environmental factors (PM<sub>2.5</sub>, PM<sub>10</sub>, temperature, and humidity) and

imported risk on the cumulative numbers of confirmed cases of COVID-19 in agglomerated and non-aggregated regions. It is of great significance for understanding the epidemiological distribution characteristics of COVID-19, providing clues on the causes of changes in the type of epidemic areas, delineating key prevention areas, and promoting the prevention and control of COVID-19.

## Materials and methods

### Sample and data

Records on COVID-19 cases between January 15 and February 25, 2020, were obtained from local health commissions on the official websites. Particulate matter  $\leq 2.5 \mu\text{m}$  (PM<sub>2.5</sub>), particulate matter  $\leq 10 \mu\text{m}$  (PM<sub>10</sub>), and average temperature and relative humidity data were obtained from the AQI platform website (<https://www.aqistudy.cn>) and the China Meteorological Date Sharing Service System (<http://cdc.cma.gov.cn>). The demographic data of each city during the study period were obtained from the statistical yearbook of each city. The population migration indexes of other regions in mainland China were obtained from Baidu Migration (<https://qianxi.baidu.com>). In our research, all COVID-19 cases were confirmed by the Chinese Center for Disease Control and Prevention with the diagnostic criteria of COVID-19(Wu et al. 2020a).

### Measures of variables

The calculation of the COVID-19 imported risk index assumes that the migrant population conforms to the population distribution in the source region, and the risk of disease is consistent with the incidence rate of the source region. The imported risk of COVID-19 caused by population migration can be expressed by the scale of migration, the number of patients in the source region, and the population. The calculation formula for the imported risk index is as follows:

$$\text{Risk}_{in,t} = \sum \left( \frac{\text{case}_{i,t}}{\text{pop}_i} \times iMI_{i,t} \times 1000 \right)$$

where Risk<sub>in,t</sub> represents the imported risk index on day *t*, case<sub>*i,t*</sub> represents the number of new cases in city *i* on day *t*, pop<sub>*i*</sub> represents the population of city *i*, and *iMI*<sub>*i,t*</sub> represents the population size index of immigration from city *i* on day *t*.

### Data analysis procedure and tests

Our study used SaTScan10.0 software to conduct scan statistics on the spatial, temporal, and space-time distributions of COVID-19. Using ArcGIS10.0 software, the number of

COVID-19 cases in 366 cities in mainland China was geocoded through management codes and matched with city-level polygons.

The analysis was performed in four stages. First, Moran's  $I$  index was used to evaluate whether a global spatial autocorrelation existed among 366 cities. A positive value of Moran's  $I$  implies a clustered distribution, whereas a negative value indicates a scattered distribution. Moran's  $I$  value close to 0 indicates a random distribution in space. Second, the discrete Poisson probability model was used to conduct Kulldorff's space-time scan statistics to explore the temporal, spatial, and spatial-temporal clusters of COVID-19. The principle of spatiotemporal scanning statistics is to use thousands of overlapping cylinders to detect spatiotemporal clusters, and each scanning window is regarded as a possible spatiotemporal cluster. The bottom of the cylinder represents the geographic area, and the height of the cylinder represents the possible duration of the outbreak. Monte Carlo calculations are performed through the Poisson model, and the log likelihood ratio and relative risk pair of each scanned cylinder are calculated. The null hypothesis of the Poisson model is that the incidence of new coronary pneumonia inside and outside the cylinder is the same. When the scan reveals that the incidence rate inside the cylinder is greater than that outside the cylinder, the invalid hypothesis can be rejected (Song and Kulldorff 2003). The spatiotemporal clusters with the largest log likelihood ratio (LLR) in the scan result are the most likely spatiotemporal clusters, whereas the other clusters are the secondary spatiotemporal clusters (Desjardins et al. 2020). Considering the human population density and the size of each city, the maximum cluster size was set to 20%. Finally, we conducted logistic regression modeling to examine the associations of the environmental factors and imported risk according to the results of the space-time clusters of COVID-19 cases. Logistic regression analysis adopts forward stepwise (conditional) for variable screening, the entry equation level is 0.05, and the elimination level is 0.10, taking particulate matters concentration and imported risk index as independent variables, and spatiotemporal aggregation as dependent variables.

## Results and discussion

In mainland China, a total of 77,658 COVID-19 cases were reported from January 15 to February 25, 2020. The average incidence rate of COVID-19 varied from 0.37/100,000 to 120.31/100,000 at the city level. The scanning results indicated that COVID-19 cases geographically differ across China. The central areas of China exhibited an increased overall risk with the central region of China showing the highest overall risk (Fig. 1).

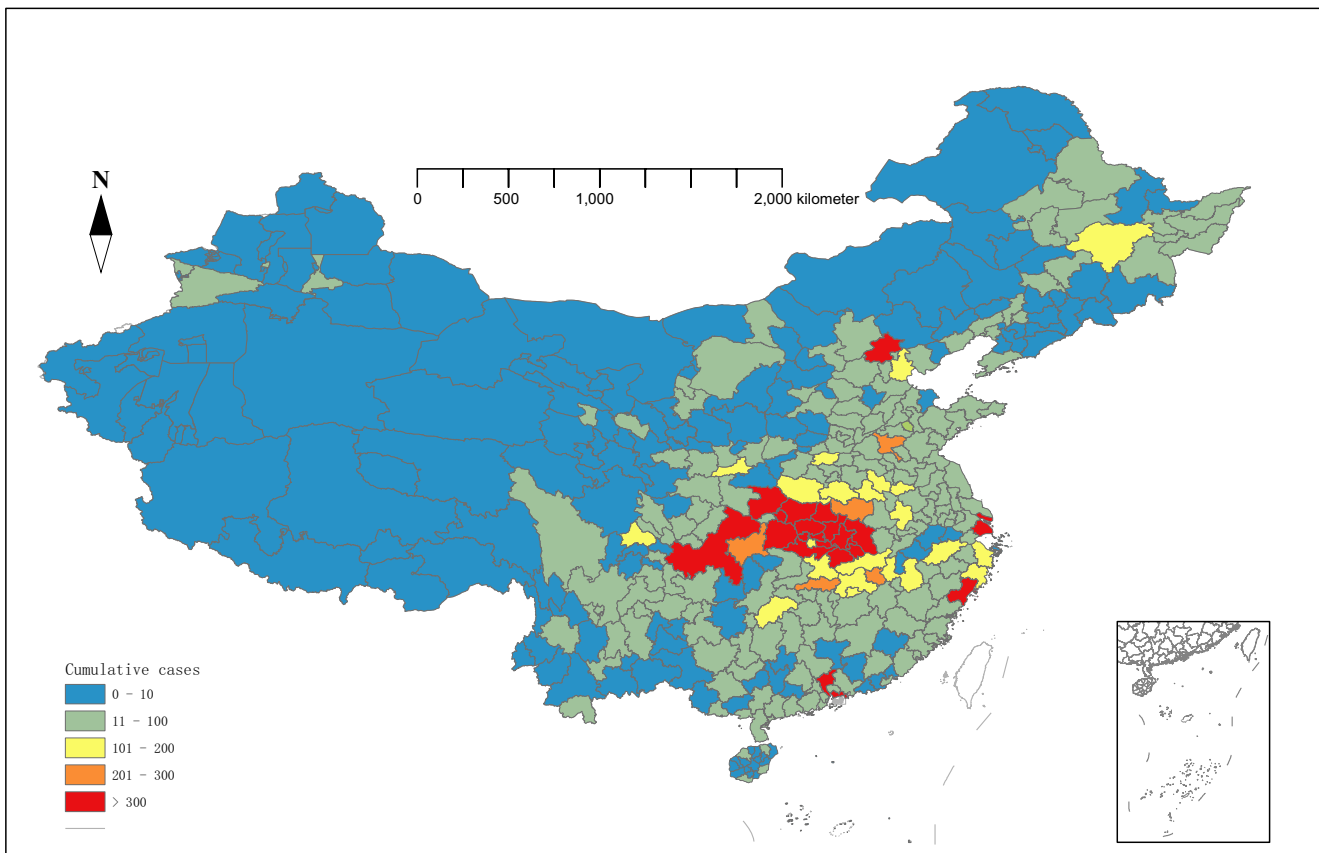
We created a Moran scatter plot and evaluated the significance of the average incidence of COVID-19 through global

spatial autocorrelation analysis. Moran's  $I$  values ranged from  $-0.004$  to  $0.453$ . The global Moran's  $I$  of the cumulative number of COVID-19 cases in 366 cities in mainland China passed the significance test (all  $P < 0.05$ ), indicating that the spatial distribution was regular from January 23 to February 25, 2020. However, the spatial distribution was not significant from January 15 to January 22, 2020 (Table 1).

The results of the spatial analysis revealed the spatial distribution characteristics of COVID-19 and identified one most likely cluster and three secondary clusters in 366 cities in mainland China (Table 2 and Fig. 2). Twenty cities were included in total, and most cities were located in Hubei Province of China. The most likely cluster included ten regions with an overall relative risk within the cluster of 96.87 (LLR = 124,907.02,  $P < 0.001$ ). The secondary cluster with a total of 8 areas was adjacent to the most likely cluster, and the relative risk was 4.51 (LLR = 3616.12,  $P < 0.001$ ). The remaining two spatial clustering regions were located in Xinyu and Wenzhou, and the relative risks were 2.49 and 1.24, respectively ( $P < 0.05$ ). In our study, most of the confirmed cases were in Wuhan, and the outbreak coincided with the Spring Festival holidays, which is an annual period of large-scale migration (Yang et al. 2020a, 2020b). Therefore, the number of newly confirmed cases in China has increased rapidly in a short period of time. In addition, the cities with more cases were all in the nearby areas adjacent to Wuhan, and there was no obvious spatial clustering in cities far away from Wuhan. This finding reflected that the population outflow from Wuhan has a great impact on the epidemic situation in the surrounding areas.

Distribution of reported COVID-19 cases based on temporal clustering indicated that COVID-19 was not random in time. One most likely cluster was identified, and the overall relative risk within the cluster was 3.10 (LLR = 8629.09,  $P < 0.001$ ). No secondary clusters were identified (Table 3). The COVID-19 incidence in our identified COVID-19 cluster was 84.3% of the total number of cases. This finding indirectly indicates that the scanning method has high sensitivity.

Spatiotemporal cluster analysis of COVID-19 cases during the study period showed that COVID-19 cases were statistically significant in space-time and identified one most likely cluster and seven secondary clusters (Table 4 and Fig. 3). The most likely cluster contained 10 cities in the province of Hubei and the high-risk time was from January 29 to February 18, 2020, which accounted for 47.83% of the total cumulative disease cases during the study period (LLR = 123,564.50, RR = 125.17,  $P < 0.001$ ). The other seven secondary clusters included 83 cities, which were mainly distributed in the central, eastern, southern, southeastern, and northeastern regions of mainland China. The overall relative risk within the cluster was statistically significant, indicating a nonrandom pattern of disease distribution ( $P < 0.001$ ).



**Fig. 1** Spatial distribution map of cumulative cases of COVID-19 in mainland China from January 15 to February 25, 2020

The results of the logistic regression model showed that environmental factors ( $PM_{2.5}$ ,  $PM_{10}$ , temperature, and humidity) and the imported risk index were strongly associated with the space-time clusters of COVID-19 cases in mainland China ( $P < 0.001$ ), and in the current model, it can explain 18.3% of the variation ( $\text{pseudo-}R^2=0.183$ ) (Table 5). Among them, temperature and  $PM_{10}$  exhibited a negative association with the space-time clusters of COVID-19 cases ( $PM_{10}$ : [OR], 0.959; 95% CI, 0.952–0.965; temperature: [OR], 0.987; 95% CI, 0.978–0.995); imported risk index, humidity, and  $PM_{2.5}$  were significant effectors on the space-time clusters of COVID-19 cases (imported risk index: [OR], 2.443; 95% CI, 2.240–2.665.  $PM_{2.5}$ : [OR], 1.035; 95%CI, 1.028–1.043; humidity: [OR], 1.025; 95%CI, 1.019–1.031).

Kulldorff’s retrospective scan statistics are one of the most effective methods using routinely collected data to explore the geographical and temporal distribution (Alemu et al. 2013; Rao et al. 2017). This method dynamically supplements the pure spatial statistical method, and is used for the prediction and detection of disease outbreaks. Prevention strategies focusing on regions of highest risk could help enhance the effectiveness of the public health interventions (Wu et al. 2011). Disease clusters were detected by using this method worldwide (Ge et al. 2016; Wang et al. 2016; Zhang et al. 2014; Zhao et al. 2013). The appropriate selection of the spatial

scanning window and temporal scanning window plays an important role in model identification in the temporal and spatial models. The temporal scanning window was usually the default setting as 50% of the whole research time, but some evidence suspected whether it was reasonable (Wang et al. 2013). A simulation study by Ma et al. found that if the window covers 50% of the population, it may be too large to cover low-risk areas, which possibly causes a high false positive rate (Ma et al. 2016). Similarly, if the scanning window was too small, it would be difficult to discover the real high-risk regions, potentially resulting in a high false negative rate (Rao et al. 2017). To identify the cluster areas with less overlap, Ge and Zhang used the irregular scan statistic to choose an appropriate window (Ge et al. 2016). Therefore, we learned from their experiences and used these findings as a reference in our research. The maximum temporal window was set as 20%, and the spatial window covered 20% of the population at risk. Finally, the identified high-risk clusters had no overlap.

Although some studies in China have explored the distribution of COVID-19 cases, these studies were limited to a certain province or region or a national study conducted only at the provincial level instead of at the city level (Yang et al. 2020a). When assessing the country, we could either obtain the clustered provinces or determine the clustering in a

**Table 1** Spatial autocorrelation analysis for the incidence of COVID-19 in mainland China from January 15 to February 25, 2020

Date	Moran's <i>I</i> index	Z-score	<i>P</i> -value
2020/1/15	0	0.000	0.001
2020/1/16	-0.004	-0.7164	0.187
2020/1/17	-0.004	-0.7164	0.187
2020/1/18	-0.004	-0.7164	0.187
2020/1/19	-0.004	-0.7164	0.187
2020/1/20	0.045	3.1692	0.086
2020/1/21	-0.005	-0.4097	0.303
2020/1/22	0.011	1.3146	0.083
2020/1/23	0.079	4.3922	0.006
2020/1/24	0.143	4.6559	0.010
2020/1/25	0.453	13.6188	0.001
2020/1/26	0.395	13.8856	0.001
2020/1/27	0.088	12.7601	0.001
2020/1/28	0.311	12.8725	0.001
2020/1/29	0.353	12.7145	0.001
2020/1/30	0.399	16.9738	0.001
2020/1/31	0.235	11.2698	0.001
2020/2/1	0.218	10.6122	0.002
2020/2/2	0.216	12.2352	0.001
2020/2/3	0.166	12.2391	0.001
2020/2/4	0.136	12.6101	0.001
2020/2/5	0.143	12.1884	0.002
2020/2/6	0.131	14.3133	0.001
2020/2/7	0.101	14.3221	0.001
2020/2/8	0.102	12.7055	0.001
2020/2/9	0.074	13.5565	0.001
2020/2/10	0.068	12.9272	0.001
2020/2/11	0.105	14.7768	0.001
2020/2/12	0.078	13.6042	0.001
2020/2/13	0.062	13.9512	0.001
2020/2/14	0.062	14.8846	0.001
2020/2/15	0.031	11.3099	0.002
2020/2/16	0.025	10.7537	0.002
2020/2/17	0.026	12.2438	0.001
2020/2/18	0.017	10.5532	0.001
2020/2/19	0.018	8.6605	0.001
2020/2/20	0.030	1.4041	0.010
2020/2/21	0.030	8.1838	0.005
2020/2/22	0.016	8.9281	0.002
2020/2/23	0.021	8.3750	0.003
2020/2/24	0.013	8.0830	0.003
2020/2/25	0.017	7.8499	0.003

specific city in the province and further analyze the impact of particulate matter and imported risk on the clustering area. In the current study, we investigated the temporal, spatial, and

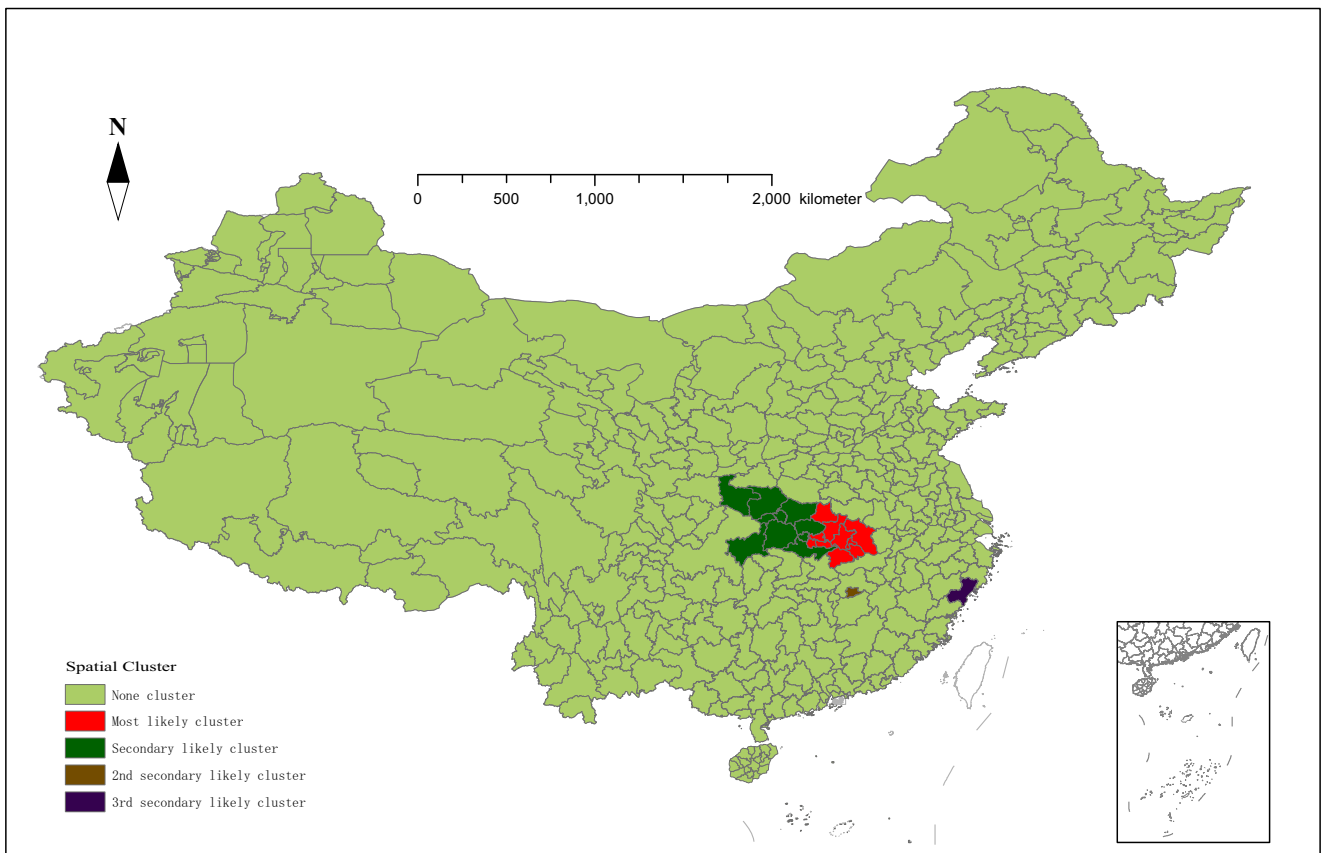
spatiotemporal distributions of COVID-19 cases and explored the potential effects of imported risk and particulate matter on spatiotemporal aggregation. We also used spatial scanning statistics to identify regions with a high prevalence of COVID-19 and clustering patterns. Our study indicated that, during the entire period from January 15 to February 25, 2020, the geographic, temporal, and spatiotemporal distribution patterns of COVID-19 cases in mainland China were not random.

Compared with separate spatial clustering analysis and temporal clustering analysis, spatiotemporal scanning statistics completely utilize the temporal and spatial information in infectious disease surveillance data and improve the timeliness and effectiveness of cluster detection (Tang et al. 2017). This analysis indicated that the COVID-19 epidemic focus had shifted from Hubei Province to other provinces of China. In addition to high population mobility, which provided favorable opportunities for the rapid spread of the virus, a number of observations point to the role of environmental factors in the survival and transmission of novel coronavirus (Bashir et al. 2020; Fattorini and Regoli 2020; Li et al. 2020). We observed that imported risk, humidity, and PM<sub>2.5</sub> were the main risk factors affecting space-time accumulation. Conversely, temperature and PM<sub>10</sub> became a protective factor against the spatiotemporal accumulation of COVID-19. The results of our research are partially consistent with the findings of Jiang et al. (2020) and Li et al. (2020), and we believe this phenomenon is related to the diameter of the particles. Studies have shown that when the diameter of particulate matter does not exceed 5 μm, it can reach type II alveolar cells, which harbor SARS-CoV-2 cell entry receptor (ACE2)(Copat et al. 2020). Although all particulate matter could facilitate viral attachment, PM<sub>2.5</sub> can deliver SARS-CoV-2 into the target cells in the alveoli whereas PM<sub>10</sub> cannot (Lee et al. 2014). This factor might also explain our research why PM<sub>2.5</sub> is positively correlated with the accumulation of new coronary pneumonia, whereas PM<sub>10</sub> is negatively correlated. In our research, temperature is a protective factor for new coronary pneumonia, and high temperature inhibits the clustering of confirmed cases of COVID-19 in Chinese cities. Some epidemiological studies also show that under low-temperature conditions, the spread of SARS-CoV-2 is more efficient and more stable (Haque and Rahman 2020). Humidity, as a risk factor, increases the possibility of temporal and spatial accumulation of new coronary pneumonia. This conclusion is similar to the research results of Jiang et al. (2020). However, many studies have inconsistent and conflicting results regarding the impact of meteorological factors on the new coronary pneumonia (Zheng et al. 2021). Therefore, further verification is necessary.

**Conclusion** Under the conditions of no imported cases from abroad, our paper identified significant spatial, temporal, and spatiotemporal clusters of COVID-19 cases at the city level in

**Table 2** Space clusters of COVID-19 cases in mainland China from January 15 to February 25, 2020

Type	Location	Cases	Expected	Relative risk	P-value
Most likely	Wuhan City, Huanggang City, Ezhou City, Xianning City, Huangshi City, Xiantao City, Qianjiang City, Tianmen City, Xiaogan City, Suizhou City	43,503	1496.82	96.87	0.001
Secondary	Shennongjia forestry district, Shiyan City, Yichang City, Xiangyang City, Jingmen City, Ankang City, Enshi City, Jingzhou City	5158	1224.16	4.51	0.001
2nd secondary	Xinyu City	130	52.22	2.49	0.001
3rd secondary	Wenzhou City	504	407.01	1.24	0.01



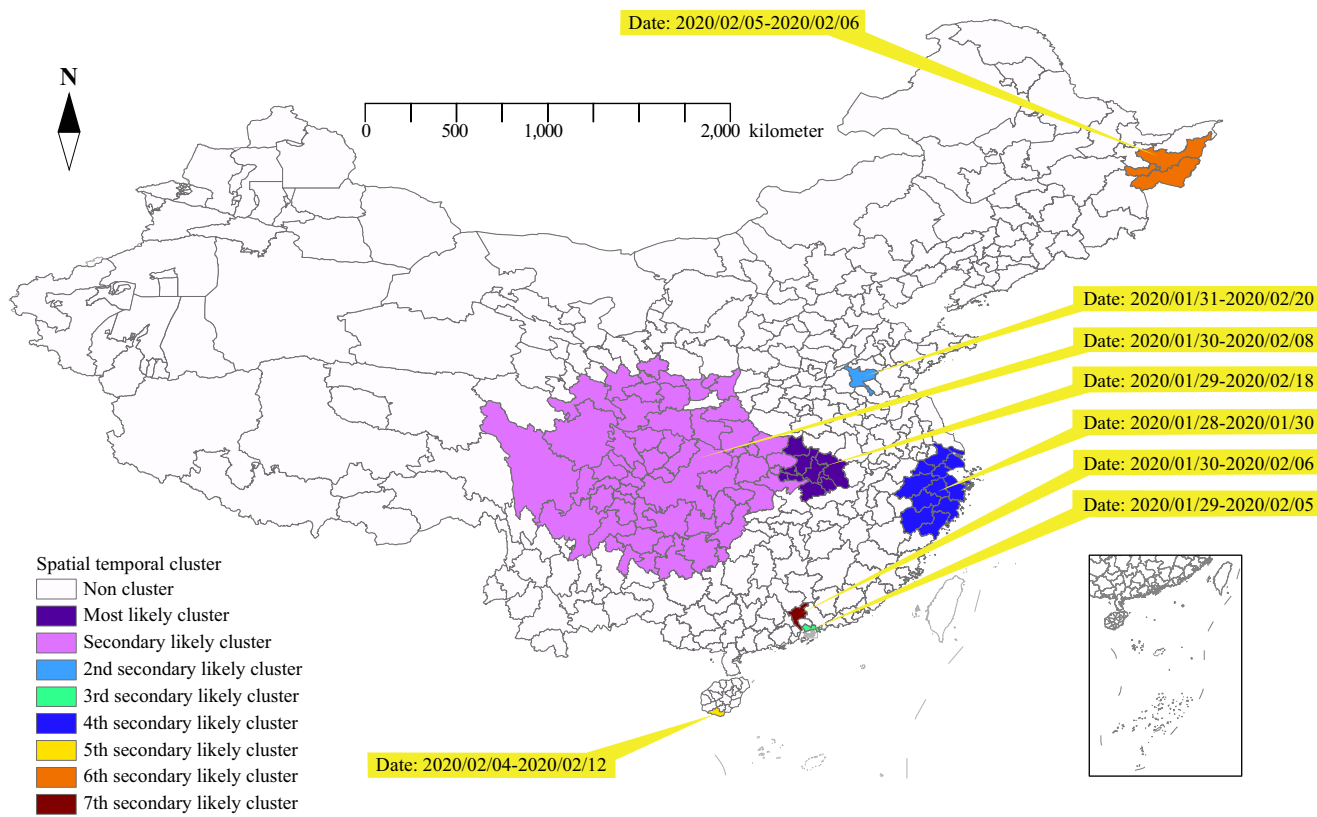
**Fig. 2** Spatial distribution of clusters of COVID-19 cases in mainland China from January 15 to February 25, 2020

**Table 3** Temporal clusters of COVID-19 cases in mainland China from January 15 to February 25, 2020

Type	Time frame	Cases	Expected	Relative risk	P-value
Most likely	2020/2/2–2020/2/9	25,919	11,715.81	3.10	0.001
Secondary	-	-	-	-	-

**Table 4** Spatiotemporal clusters of COVID-19 cases in mainland China from January 15 to February 25, 2020

Type	Time frame	Coordinates/radius	N	Cases	Expected	Relative risk	P-value
Most likely	2020/1/29–2020/2/18	(30.52 N, 114.31 E) 160.56 km	10	37309	748.41	125.17	0.001
Secondary	2020/1/30–2020/2/8	(30.78 N, 106.08 E) 597.91 km	54	3957	2385.55	1.70	0.001
2nd secondary	2020/1/31–2020/2/20	(35.38 N, 116.59 E) 0 km	1	201	8.74	23.06	0.001
3rd secondary	2020/1/29–2020/2/5	(22.62 N, 114.07 E) /0 km	1	251	109.18	2.30	0.001
4th secondary	2020/1/28–2020/1/30	(29.86 N, 121.56 E) 223.82 km	12	439	283.25	1.55	0.001
5th secondary	2020/2/4–2020/2/12	(18.25 N, 109.50 E) 0 km	1	36	7.30	4.94	0.001
6th secondary	2020/2/5–2020/2/6	(45.77 N, 131.02 E) 98.47 km	3	38	8.20	4.64	0.001
7th secondary	2020/1/30–2020/2/6	(23.16 N, 113.23 E) 0 km	1	205	124.92	1.64	0.001



**Fig. 3** Spatiotemporal distribution of clusters of COVID-19 cases in mainland China from January 15 to February 25, 2020

**Table 5** The association between the space-time clusters of COVID-19 cases and environmental factors and imported risk index

Variables	Regression coefficient	Odds ratio	95% confidence interval	P-value	Pseudo-R <sup>2</sup>
Imported risk index	0.893	2.443	2.240–2.665	0.001	0.183
PM <sub>2.5</sub>	0.035	1.035	1.028–1.043	0.001	
PM <sub>10</sub>	−0.042	0.959	0.952–0.965	0.001	
Humidity	0.025	1.025	1.019–1.031	0.001	
Temperature	−0.013	0.987	0.978–0.995	0.002	

mainland China. In addition, the impacts of environmental factors and imported risks on spatiotemporal clustering were also revealed. Hubei province and its surrounding cities were the most likely clustering areas with the migration of population and the impact of climate factors. A significant correlation is noted between imported risk, humidity, and PM<sub>2.5</sub> and the spatiotemporal accumulation of COVID-19, whereas temperature and PM<sub>10</sub> are negatively correlated with the spatiotemporal clustering of COVID-19.

The cluster analysis in this study is an ecological study, the quality of data over time and from different places may be quite variable, and the ecological fallacy is usually interpreted as a major weakness. Therefore, these results cannot be extrapolated to the individual level. Another weakness is the limitations of some traditional statistical methods in our research. The occurrence and development of new coronary pneumonia are complicated, and there are often correlations between observation indicators. When constructing the logistic linear regression equation, it appears as collinearity between variables, which may cause some important indicators to be screened out. However, due to its strong versatility, high fitting accuracy, and wide application value, the results are still convincing. And cluster analysis does help to visualize epidemiological data and detect and evaluate hot spots or clusters. The results may improve disease surveillance and effectively manage control plan resources.

In addition, the pandemic of new coronary pneumonia and the mutation of viral factors have caused many socio-economic problems (Coccia 2021c). Because it is a new pathogen that is highly infectious and concealed, humans generally lack immunity to it. Therefore, the primary response strategy in mainland China is to adopt physical isolation, and nucleic acid testing was used to detect and strip the source of infection early for isolation and centralized treatment. These measures not only prevent the virus in the discovered patient from spreading to the uninfected person, but also prevent the virus discharged by the asymptomatic person who has not been found to be positive with the virus to enter the next susceptible person and be inactivated in the environment. The new coronavirus will no longer exist in the area until after one or at most two incubation periods. Secondly, with the successful development of vaccines, it is necessary to do our best to complete the whole course of COVID-19 vaccination in the largest range and in the shortest time, so as to achieve the effect of herd immunity. Moreover, because cases continue to appear in our country, great uncertainties remain in the late stage of the epidemic. The incidence of COVID-19 cases should be closely monitored on a continuous basis, and targeted prevention and control strategies should be developed to effectively control the epidemic of COVID-19.

**Author contribution** SY designed and drafted the manuscript. SY, CL, and PG participated in the data collection. ZF, DH, and WW participated in the data analysis. SA and WW made critical revision to the manuscript. The final manuscript has been read and approved by all the authors.

**Funding** This study was supported by the National Natural Science Foundation of China (Grant Nos. 81202254 and 71974199).

**Data availability** The data were obtained from local health commissions on the official websites, the statistical yearbook of each city, AQI platform website (<https://www.aqistudy.cn>), the China Meteorological Data Sharing Service System (<http://cdc.cma.gov.cn>), and Baidu Migration (<https://qianxi.baidu.com>).

## Declarations

**Ethics approval and consent to participate** Not applicable

**Consent for publication** Not applicable

**Competing interests** The authors declare no competing interests.

## References

- Alemu K, Worku A, Berhane Y (2013) Malaria infection has spatial, temporal, and spatiotemporal heterogeneity in unstable malaria transmission areas in northwest Ethiopia. *PLoS One* 8(11):e79966. <https://doi.org/10.1371/journal.pone.0079966>
- Bashir MF, Ma BJ, Bilal, Komal B, Bashir MA, Farooq TH et al (2020) Correlation between environmental pollution indicators and COVID-19 pandemic: a brief study in Californian context. *Environ Res* 187:109652. <https://doi.org/10.1016/j.envres.2020.109652>
- Chan JF, Yuan S, Kok KH, To KK, Chu H, Yang J et al (2020) A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* 395(10223):514–523. [https://doi.org/10.1016/s0140-6736\(20\)30154-9](https://doi.org/10.1016/s0140-6736(20)30154-9)
- Chen G, Tao J, Wang J, Dong M, Li X, Sun X, Cheng S, Fan J, Ye Y, Xiao J, Hu J, He G, Sun J, Lu J, Guo L, Li X, Rong Z, Zeng W, Zhou H et al (2021) Reduction of air pollutants and associated mortality during and after the COVID-19 lockdown in China: impacts and implications. *Environ Res* 200:111457. <https://doi.org/10.1016/j.envres.2021.111457>
- Chen K, Wang M, Huang C, Kinney PL, Anastas PT (2020) Air pollution reduction and mortality benefit during the COVID-19 outbreak in China. *Lancet Planet Health* 4(6):e210–e212. [https://doi.org/10.1016/s2542-5196\(20\)30107-8](https://doi.org/10.1016/s2542-5196(20)30107-8)
- Coccia M (2020a) Factors determining the diffusion of COVID-19 and suggested strategy to prevent future accelerated viral infectivity similar to COVID. *Sci Total Environ* 729:138474. <https://doi.org/10.1016/j.scitotenv.2020.138474>
- Coccia M (2020b) An index to quantify environmental risk of exposure to future epidemics of the COVID-19 and similar viral agents: theory and practice. *Environ Res* 191:110155. <https://doi.org/10.1016/j.envres.2020.110155>
- Coccia M (2021) The effects of atmospheric stability with low wind speed and of air pollution on the accelerated transmission dynamics of COVID-19. *International Journal of Environmental Studies* 78(1):1–27. <https://doi.org/10.1080/00207233.2020.1802937>
- Coccia M (2021a) Effects of the spread of COVID-19 on public health of polluted cities: results of the first wave for explaining the déjà vu in



- the second wave of COVID-19 pandemic and epidemics of future vital agents. *Environ Sci Pollut Res Int* 28(15):19147–19154. <https://doi.org/10.1007/s11356-020-11662-7>
- Coccia M (2021b) How do low wind speeds and high levels of air pollution support the spread of COVID-19? *Atmos Pollut Res* 12(1): 437–445. <https://doi.org/10.1016/j.apr.2020.10.002>
- Coccia M (2021c) Preparedness of countries to face COVID-19 pandemic crisis: strategic positioning and underlying structural factors to support strategies of prevention of pandemic threats. *Environ Res* 203:111678. <https://doi.org/10.1016/j.envres.2021.111678>
- Copat C, Cristaldi A, Fiore M, Grasso A, Zuccarello P, Signorelli SS, Conti GO, Ferrante M (2020) The role of air pollution (PM and NO<sub>2</sub>) in COVID-19 spread and lethality: a systematic review. *Environ Res* 191:110129. <https://doi.org/10.1016/j.envres.2020.110129>
- Desjardins MR, Hohl A, Delmelle EM (2020) Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: detecting and evaluating emerging clusters. *Appl Geogr* 118:102202. <https://doi.org/10.1016/j.apgeog.2020.102202>
- Du H, Li J, Wang Z, Yang W, Chen X, Wei Y (2021) Sources of PM<sub>2.5</sub> and its responses to emission reduction strategies in the Central Plains Economic Region in China: implications for the impacts of COVID-19. *Environ Pollut* 288:117783. <https://doi.org/10.1016/j.envpol.2021.117783>
- Fang LQ, Zhao WJ, de Vlas SJ, Zhang WY, Liang S, Looma CW et al (2009) Spatiotemporal dynamics of hemorrhagic fever with renal syndrome, Beijing, People's Republic of China. *Emerg Infect Dis* 15(12):2043–2045. <https://doi.org/10.3201/eid1512.081078>
- Fattorini D, Regoli F (2020) Role of the chronic air pollution levels in the COVID-19 outbreak risk in Italy. *Environ Pollut* 264:114732. <https://doi.org/10.1016/j.envpol.2020.114732>
- Fosgate GT, Carpenter TE, Chomel BB, Case JT, DeBess EE, Reilly KF (2002) Time-space clustering of human brucellosis, California, 1973–1992. *Emerg Infect Dis* 8(7):672–678. <https://doi.org/10.3201/eid0807.010351>
- Ge E, Zhang X, Wang X, Wei X (2016) Spatial and temporal analysis of tuberculosis in Zhejiang Province, China, 2009–2012. *Infect Dis Poverty* 5:11. <https://doi.org/10.1186/s40249-016-0104-2>
- Haque SE, Rahman M (2020) Association between temperature, humidity, and COVID-19 outbreaks in Bangladesh. *Environ Sci Policy* 114:253–255. <https://doi.org/10.1016/j.envsci.2020.08.012> <https://covid19.who.int/table>
- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M et al (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395(10223):497–506. [https://doi.org/10.1016/s0140-6736\(20\)30183-5](https://doi.org/10.1016/s0140-6736(20)30183-5)
- Irfan, M., Ikram, M., Ahmad, M., Wu, H., & Hao, Y. (2021). Does temperature matter for COVID-19 transmissibility? Evidence across Pakistani provinces. *Environ Sci Pollut Res Int*, 1–15. doi:<https://doi.org/10.1007/s11356-021-14875-6>
- Islam N, Bukhari Q, Jameel Y, Shabnam S, Erzurumluoglu AM, Siddique MA, Massaro JM, D'Agostino RB Sr (2021) COVID-19 and climatic factors: a global analysis. *Environ Res* 193:110355. <https://doi.org/10.1016/j.envres.2020.110355>
- Jiang Y, Wu XJ, Guan YJ (2020) Effect of ambient air pollutants and meteorological variables on COVID-19 incidence. *Infect Control Hosp Epidemiol* 41(9):1011–1015. <https://doi.org/10.1017/ice.2020.222>
- Konstantinou G, Padellini T, Bennett J, Davies B, Ezzati M, Blangiardo M (2021) Long-term exposure to air-pollution and COVID-19 mortality in England: a hierarchical spatial analysis. *Environ Int* 146:106316. <https://doi.org/10.1016/j.envint.2020.106316>
- Lai WT, Chen CH, Hung H, Chen RB, Shete S, Wu CC (2018) Recognizing spatial and temporal clustering patterns of dengue outbreaks in Taiwan. *BMC Infect Dis* 18(1):256. <https://doi.org/10.1186/s12879-018-3159-9>
- Lee GI, Saravia J, You D, Shrestha B, Jaligama S, Hebert VY, Dugas TR, Cormier SA (2014) Exposure to combustion generated environmentally persistent free radicals enhances severity of influenza virus infection. *Part Fibre Toxicol* 11:57. <https://doi.org/10.1186/s12989-014-0057-1>
- Li H, Xu XL, Dai DW, Huang ZY, Ma Z, Guan YJ (2020) Air pollution and temperature are associated with increased COVID-19 incidence: a time series study. *Int J Infect Dis* 97:278–282. <https://doi.org/10.1016/j.ijid.2020.05.076>
- Lu H, Stratton CW, Tang YW (2020) Outbreak of pneumonia of unknown etiology in Wuhan, China: the mystery and the miracle. *J Med Virol* 92(4):401–402. <https://doi.org/10.1002/jmv.25678>
- Ma Y, Yin F, Zhang T, Zhou XA, Li X (2016) Selection of the maximum spatial cluster size of the spatial scan statistic by using the maximum clustering set-proportion statistic. *PLoS One* 11(1):e0147918. <https://doi.org/10.1371/journal.pone.0147918>
- Manisalidis I, Stavropoulou E, Stavropoulos A, Bezirtzoglou E (2020) Environmental and health impacts of air pollution: a review. *Front Public Health* 8:14. <https://doi.org/10.3389/fpubh.2020.00014>
- Rahimi NR, Fouladi-Fard R, Aali R, Shahryari A, Rezaali M, Ghafouri Y, Ghalhari MR, Asadi-Ghalhari M, Farzinnia B, Conti Gea O, Fiore M (2021) Bidirectional association between COVID-19 and the environment: a systematic review. *Environ Res* 194:110692. <https://doi.org/10.1016/j.envres.2020.110692>
- Rao H, Shi X, Zhang X (2017) Using the Kulldorff's scan statistical analysis to detect spatio-temporal clusters of tuberculosis in Qinghai Province, China, 2009–2016. *BMC Infect Dis* 17(1):578. <https://doi.org/10.1186/s12879-017-2643-y>
- Rosario DKA, Mutz YS, Bernardes PC, Conte-Junior CA (2020) Relationship between COVID-19 and weather: case study in a tropical country. *Int J Hyg Environ Health* 229:113587. <https://doi.org/10.1016/j.ijheh.2020.113587>
- Sarkodie SA, Owusu PA (2020) Impact of meteorological factors on COVID-19 pandemic: evidence from top 20 countries with confirmed cases. *Environ Res* 191:110101. <https://doi.org/10.1016/j.envres.2020.110101>
- Shen L, Zhao T, Wang H, Liu J, Bai Y, Kong S, Zheng H, Zhu Y, Shu Z (2021) Importance of meteorology in air pollution events during the city lockdown for COVID-19 in Hubei Province, Central China. *Sci Total Environ* 754:142227. <https://doi.org/10.1016/j.scitotenv.2020.142227>
- Song C, Kulldorff M (2003) Power evaluation of disease clustering tests. *Int J Health Geogr* 2(1):9. <https://doi.org/10.1186/1476-072x-2-9>
- Srivastava A (2021) COVID-19 and air pollution and meteorology—an intricate relationship: a review. *Chemosphere* 263:128297. <https://doi.org/10.1016/j.chemosphere.2020.128297>
- Tang X, Geater A, McNeil E, Deng Q, Dong G, Zhong G (2017) Spatial, temporal and spatio-temporal clusters of measles incidence at the county level in Guangxi, China during 2004–2014: flexibly shaped scan statistics. *BMC Infect Dis* 17(1):243. <https://doi.org/10.1186/s12879-017-2357-1>
- Tian F, Liu X, Chao Q, Qian ZM, Zhang S, Qi L et al (2021) Ambient air pollution and low temperature associated with case fatality of COVID-19: a nationwide retrospective cohort study in China. *Innovation (N Y)* 2(3):100139. <https://doi.org/10.1016/j.xinn.2021.100139>
- Tian L, Liang F, Xu M, Jia L, Pan X, Clements ACA (2018) Spatio-temporal analysis of the relationship between meteorological factors and hand-foot-mouth disease in Beijing, China. *BMC Infect Dis* 18(1):158. <https://doi.org/10.1186/s12879-018-3071-3>
- Vlachogiannis DM, Xu Y, Jin L, González MC (2021) Correlation networks of air particulate matter (PM<sub>2.5</sub>): a comparative study. *Appl Netw Sci* 6(1):32. <https://doi.org/10.1007/s41109-021-00373-8>

- Wang LY, Zhang WY, Ding F, Hu WB, Soares Magalhaes RJ, Sun HL, Li YX, Zou W, Wang Y, Liu QY, Li SL, Yin WW, Huang LY, Clements ACA, Bi P, Li CY (2013) Spatiotemporal patterns of Japanese encephalitis in China, 2002-2010. *PLoS Negl Trop Dis* 7(6):e2285. <https://doi.org/10.1371/journal.pntd.0002285>
- Wang WL, Wang HJ, Deng Y, Song T, Lan JM, Wu GZ, Ke CW, Tan WJ (2016) Serological study of an imported case of Middle East respiratory syndrome and his close contacts in China, 2015. *Biomed Environ Sci* 29(3):219–223. <https://doi.org/10.3967/bes2016.027>
- Wu JT, Leung K, Leung GM (2020a) Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* 395(10225):689–697. [https://doi.org/10.1016/s0140-6736\(20\)30260-9](https://doi.org/10.1016/s0140-6736(20)30260-9)
- Wu S, Ni Y, Li H, Pan L, Yang D, Baccarelli AA, Deng F, Chen Y, Shima M, Guo X (2016) Short-term exposure to high ambient air pollution increases airway inflammation and respiratory symptoms in chronic obstructive pulmonary disease patients in Beijing, China. *Environ Int* 94:76–82. <https://doi.org/10.1016/j.envint.2016.05.004>
- Wu W, Guo J, Guan P, Sun Y, Zhou B (2011) Clusters of spatial, temporal, and space-time distribution of hemorrhagic fever with renal syndrome in Liaoning Province, Northeastern China. *BMC Infect Dis* 11:229. <https://doi.org/10.1186/1471-2334-11-229>
- Yan-Feng, G., Lei, L., Zhi-Hong, L., Jun, S. G., & Jian-Gang, Z. (2019). [Spatial - temporal distribution of malaria in Jiangxi Province from 1950 to 2017]. *Zhongguo Xue Xi Chong Bing Fang Zhi Za Zhi*, 31(4), 414-417. doi:10.16250/j.32.1374.2018237
- Yang W, Deng M, Li C, Huang J (2020a) Spatio-temporal patterns of the 2019-nCoV epidemic at the county level in Hubei Province, China. *Int J Environ Res Public Health* 17(7). <https://doi.org/10.3390/ijerph17072563>
- Yang, Z., Zeng, Z., Wang, K., Wong, S. S., Liang, W., Zanin, M., . . . He, J. (2020b). Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J Thorac Dis*, 12(3), 165-174. doi:10.21037/jtd.2020.02.64
- Yao Y, Pan J, Wang W, Liu Z, Kan H, Qiu Y, Meng X, Wang W (2020) Association of particulate matter pollution and case fatality rate of COVID-19 in 49 Chinese cities. *Sci Total Environ* 741:140396. <https://doi.org/10.1016/j.scitotenv.2020.140396>
- Yu-Wan, H., Feng-Hua, G., Jing-Bo, X., Jun-Fang, X., Qiang, W., Yun, Z., . . . Shi-Zhu, L. (2019). [Spatial - temporal clustering analysis of schistosomiasis transmission risk in Yunnan Province from 2004 to 2015]. *Zhongguo Xue Xi Chong Bing Fang Zhi Za Zhi*, 31(3), 269-274. doi:10.16250/j.32.1374.2019138
- Zhang WY, Wang LY, Liu YX, Yin WW, Hu WB, Magalhaes RJ et al (2014) Spatiotemporal transmission dynamics of hemorrhagic fever with renal syndrome in China, 2005-2012. *PLoS Negl Trop Dis* 8(11):e3344. <https://doi.org/10.1371/journal.pntd.0003344>
- Zhao F, Cheng S, He G, Huang F, Zhang H, Xu B, Murimwa TC, Cheng J, Hu D, Wang L (2013) Space-time clustering characteristics of tuberculosis in China, 2005-2011. *PLoS One* 8(12):e83605. <https://doi.org/10.1371/journal.pone.0083605>
- Zheng, H. L., Guo, Z. L., Wang, M. L., Yang, C., An, S. Y., & Wu, W. (2021). Effects of climate variables on the transmission of COVID-19: a systematic review of 62 ecological studies. *Environ Sci Pollut Res Int*, 1-18. doi:<https://doi.org/10.1007/s11356-021-15929-5>

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