Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Do spatiotemporal units matter for exploring the microgeographies of epidemics?

Sui Zhang, Minghao Wang, Zhao Yang, Baolei Zhang* 

College of Geography and Environment, Shandong Normal University, Jinan, 250014, China

A R T I C L E   I N F O

Keywords: 
Spatiotemporal units 
Modifiable areal unit problem 
Spatial epidemiology 
Microgeography 
COVID-19

A B S T R A C T

From the onset of the COVID-19 pandemic in 2020, studies on the microgeographies of epidemics have surged. However, studies have neglected the significant impact of multiple spatiotemporal units, such as report timestamps and spatial scales. This study examines three cities with localized COVID-19 resurgence after the first wave of the pandemic in mainland China to estimate the differential impact of spatiotemporal unit on exploring the influencing factors of epidemic spread at the microscale. The quantitative analysis results suggest that future spatial epidemiology research should give greater attention to the “symptom onset” timestamp instead of only the “confirmed” data and that “spatial transmission” should not be confused with “spatial sprawling” of epidemics, which can greatly reduce comparability between epidemiology studies. This research also highlights the importance of considering the modifiable areal unit problem (MAUP) and the uncertain geographic context problem (UGCoP) in future studies.

1. Introduction

The novel coronavirus disease 2019 (COVID-19), which was first discovered in Wuhan, China, in December 2019, has swept the world with high infectivity and mortality (Wu, Zhao, et al., 2020). In the past two years, the global public health crisis has penetrated almost all fields of academia, and more than 250,000 related papers have been published according to the web of science (WOS) database (www.webofscience.com). For geographers, COVID-19’s innate dual attributes of space and time have given them a vast role to play (Rosenkrantz, Schuurman, Bell, & Amram, 2021). Based on the temporal and spatial attributes of the pandemic, geographers have made fruitful achievements in establishing statistical analysis systems and developing abundant quantitative analytic methods for spatiotemporal epidemiology. Both efforts have made great contributions not only to the prevention and control of the pandemic but also to the smooth operation and sustainable development of society in the post-pandemic era (Guan, Wang, et al., 2020; Kraemer et al., 2020; Lai et al., 2015; Lai et al., 2017).

Because of the essential role geography plays in contemporary epidemiological research, it is undoubtedly of great significance to standardize the related research in processes and interpretations of results, build a more communicable research network, enhance the comparative value among studies and make the obtained results easier to systematically summarize (Franch-Pardo, Napoletano, Rosete-Verges, & Billa, 2020). The selection of spatiotemporal units has always been an unavoidable basic step in the microgeographic research of epidemics. First, the data about infected cases usually have multiple attributes in the time dimension, and the three attributes of “confirmed (CT)”, “quarantined/hospitalized (QT)” and “symptom onset (ST)” are the most frequently discussed (Chan et al., 2020; Chen et al., 2020; Huang et al., 2020). Since there is plenty of evidence to prove that there is a certain incubation period between viral infection and symptom onset in most epidemics and the efficiency and priority of nonpharmaceutical interventions vary spatiotemporally, these three attributes of cases often show considerable differentiation (Bertozzi, Franco, Mohler, Short, & Sledge, 2020; Bertuzzo et al., 2020; Chen et al., 2013; Chan et al., 1997; Gao et al., 2020; He, Deng, & Li, 2020; Pantaleo et al., 1993; Zhai, Liu, Fu, & Peng, 2021). Second, when locating the infection cases spatially, the scales of the spatial units become problems that cannot be ignored. Generally, microgeographers are accustomed to analysing geographic information of cases with three scales: individual (IS), cluster (CS) and area (AS). When a case is reported, the department concerned usually locates it at home or an activity place, mostly a community or a village as its smallest spatial scale, which is represented by points on a map.
When researchers try to analyse the spatial characteristics of infected cases at the individual scale, these communities/villages are regarded as a single point with multiple overlapping points (individuals) (Samphutthanon, Tripathi, Ninsawat, & Duboz, 2014; Tanser, Barnighausen, Cooke, & Newell, 2009; Wu, Yan, et al., 2020); there is also a part of the research that analyses the specific location (coordinates) for the places where cases cluster (the communities/villages mentioned above), and ignores the heterogeneity of cases within each cluster (Liu, Qin, Xie, & Zhang, 2020). To make a more official portrait of regional epidemic characteristics and assess transmission risks, many relevant studies and government documents take neighbourhood-level administrative polygon as the spatial scale for case statistics (areas), which makes the spatial pattern of cases less detailed than the first two spatial scales (Briz-Redon & Serrano-Aroca, 2020; Gatrell & Bailey, 1996; Lee, Robertson, & Marques, 2021; Yang, Deng, Li, & Huang, 2020). However, although researchers will comprehensively consider their objectives when choosing the spatiotemporal unit, the differentiation mentioned above and the following consequences are rarely quantitatively evaluated and discussed in depth, which makes it difficult for researchers to systematically and comprehensively compare and review these epidemiological studies with different spatiotemporal units, thus greatly hindering the systematization and theorization of spatiotemporal epidemiological research, for example, the ongoing pandemic COVID-19 (Pearce, Vandenbroucke, VanderWeele, & Greenland, 2020).

In this study, we selected three cities with local COVID-19 resurgence after the first wave of the pandemic in mainland China. The above-mentioned spatiotemporal units were selected to analyse the micro-geographic characteristics of the epidemic and the impact on the assessment of factors that influenced epidemic transmission. The main aims of research included the following: (1) to summarize the pattern of localized epidemic when selecting different spatiotemporal units, (2) to estimate the influence of place-based factors, “source-case” distance and spatial spillover effect on the development of localized epidemic and its temporal dynamics, (3) systematically compare the differentiation of analysis results when selecting different spatio-temporal units and discuss the application norms and scenarios.

2. Materials and methodology

2.1. Empirical cases

Three cases of localized epidemic resurgence in the year after the end of the first nationwide wave of the pandemic begun in Wuhan in January 2020 were selected in this study. These three local outbreaks have a common characteristic, that is, they have a single place-based epidemic source, and all cases are directly or indirectly associated with the local source (Liu, Yang, et al., 2020; Pang et al., 2020; Wang, Zhang, & Cai, 2021; Zhang et al., 2020): the Beijing outbreak originated from “Xinfadi Market (XFM)” in June to July 2020, the Dalian outbreak originated from “Kaiyai Seafood Company (KSF)” in July to August 2020 and the Tonghua outbreak originated from “Yuansheng Quality Living Square (YQL)” in January to February 2021. Since the Health Committee of Tonghua City stopped publishing the microgeographic data of the follow-up asymptomatic to confirmed cases from January 30th, the research limited by January 29th. To further explore the spread of COVID-19, which is mainly decided by population mobility, we took the functional urban area (FUA) delineation results using massive Didi ride-hailing records as the local study areas (Ma and Long, 2020). Considering the inequality of administrative areas and the complexity of quadrilateral grids in dealing with the problem of grid adjacency, we divided the FUAs into multiple hexagonal grids with a length of 500 m referring to the idea of community grid epidemic control and prevention (Birch, Oom, & Beecham, 2007; Li & Gao, 2020; Ling & Wen, 2020).

2.2. Data source and variable selection

The data used in this paper included case data of COVID-19 and point of interest (POI) data. The COVID-19 case data used in this study were obtained from the Municipal Health Commission or Center for Disease Control and Prevention of the cities involved in this research (Beijing Municipal Health Commission: www.beijing.gov.cn, Beijing Center for Disease Control and Prevention: www.bjcdc.org; Health Commission of Dalian: hcod.dl.gov.cn; Health Committee of TongHua City: www.tonghua.gov.cn/wjwj). The POI data were obtained from Amap (www. amap.com) and rectified deviations manually. Catering places, residential areas, shopping places, public service facilities, and health-care facilities were selected as the gathering place factors, and all these variables had variance inflation factor (VIF) values of less than 7.5 with different spatial scales, which can be accepted by models with strong interpretability.

2.3. Methodology

2.3.1. Temporal characteristics recognition

The exponential modified Gaussian function (EMG), which is usually used in the peak fitting calculation of chromatographic peaks, was used to identify the temporal trends and the turning point of the local COVID-19 outbreak (Grushka, 1972). The days before and after the turning point of the EMG were divided into the spread duration (SD) and the decay duration (DD). Given that the value corresponding to the turning point of the fitting curve is usually not an integer and that case characteristics near the turning point are similar, an extra day of cases was taken for both durations in the analysis.

2.3.2. Spatial pattern analysis

Standard Deviational Ellipse (SDE) describes the geographical distribution trend by summarizing the dispersion and directivity of observation samples (Lefever, 1926; Vull, 1971). It was used to make a basic judgement on the direction of the spread and the characteristics of the epidemic. Ripley’s K function, which is also a suitable local second-order point pattern analysis quantitative approach, was used to evaluate the distribution characteristics of features (clustered, random, or dispersed) on multiple spatial scales (Boots & Getis, 1988; Getis, 1984; Wiegand & Moloney, 2004). After the L(d) transformation, the formula is as follows:

\[
L(d) = \sqrt[1/n]{\frac{\sum_{i=1}^{n} \sum_{j=1, i \neq j}^{n} k_{ij}}{\pi n(n-1)}}
\]

where A is the total area of the analysis region; d is the search radius; n is the total number of features; and k_{ij} are binary piecewise functions added to set the distance threshold. By analysing DfHK, the differentiation between the L(d) computed and the corresponding d, the spatial autocorrelation pattern of feature distribution can be quantitatively measured. The above spatial analysis and visualization were executed with GeoDa and ArcGIS 10.7 software.

2.3.3. Spatial zero-inflated negative binomial models

The processing of count data is usually involved in the study of population flow-oriented events such as epidemics. Previous researchers prefer Poisson regression or negative binomial regression models to replace traditional ordinary least squares (OLS) regression (Flowerdew & Atkinson, 1982; Flowerdew & Boyle, 1995). However, especially in the case of this clustering epidemic, the probability of zero value in the model is often underestimated, which is beyond the predictive ability of general counting models such as Poisson and negative binomial regression. Because there are too many zeros in the counting data and the same zeros represent different situations, the counting data often show great variation. The estimation process of the zero-inflated model is composed of a logit regression and a Poisson regression or negative
binomial regression, so the regression is divided into two independent processes to solve the problem of zero value from different sources (Burger, Van Oort, & Linders, 2009; Lambert, 1992). Taking the zero-inflated negative binomial model with more widespread utilization as an example, the functions are as follows:

$$\Pr(n_i = 0) = \psi_k + (1 - \psi_k) \left( \frac{a^{-1}}{a^{-1} + \lambda_0} \right)^{a^{-1}}$$  \hspace{1cm} (4.1)

$$\Pr(n_i = k) = (1 - \psi_k) \frac{\Gamma(k + a^{-1})}{k!a^{-1}} \left( \frac{a^{-1}}{a^{-1} + \lambda_0} \right)^{a^{-1}} \left( \frac{\lambda_0}{a^{-1} + \lambda_0} \right)^k$$  \hspace{1cm} (4.2)

The conditional mean $\lambda_0$ is related to the exponential function of the regression explanatory variable:

$$\lambda_0 = \exp(\beta_0 + \beta_1 \ln P_i + \beta_2 \ln P_j - \beta_3 \ln d_{ij})$$  \hspace{1cm} (5)

By incorporating the spatial lag term of the variables into the negative binomial regression part of the model, formula (5) is transformed into the following formula:

$$\lambda_0 = \exp(\beta_0 + \beta_1 \ln P_i + \beta_2 \ln P_j + \delta W_{ij} \ln \lambda_0 - \beta_3 \ln d_{ij})$$  \hspace{1cm} (6)

When there is no excess zero in the dependent variable, the model will run under the classical negative binomial regression framework to ensure the efficiency of the model in estimating the effects of independent variables. Considering the frequent occurrence of zero values in the gathering place variables, natural logarithm operations cannot be performed. To ensure the successful calculation of the model and the reliability of the results, the independent variables are transformed by adding a fixed amount of $10^{-4}$ in this research (Table 1).

3. Why do spatiotemporal units matter?

Based on the time series analysis of the number of new cases per day in COVID-19, the temporal characteristics and segmentation of the epidemic were visualized (Fig. 1). The morphological characteristics of the fitting curves showed the trends of epidemic development with different report timestamps. Compared with the Gaussian fitting curve, which was almost similar to the characteristic of the curve of “symptom onset”, the fitting curves for “confirmed” and “quarantined” showed a strong right-sided state. This phenomenon indicated that the incidence of the COVID-19 epidemic followed a normal distribution, and its temporal characteristics were difficult to artificially change at the present stage. Through quantitative comparison of the occurrence dates of the turning points, the turning points of “confirmed” and “quarantined” were approximately 2–3 days ahead of the arrival of the turning point of “symptom onset”, respectively, further confirming the importance of strong investigation, quarantine and lockdown measures for controlling the transmission speed of local epidemics in a timely manner.

To understand the importance of spatial scale selection in epidemiological research, the standardized spatial patterns of localized outbreaks with three spatial scales were demonstrated (Fig. 2). The standardized spatial pattern reports the strong spatial heterogeneity of epidemic statistics results when adopting these three spatial scales. It is worth noting that this spatial differentiation is smaller in areas far from the epidemic source, and larger in areas close to the epidemic source, which are also the high incidence areas. If the three kinds of spatial scales were confused, there would be a serious deviation in describing the spatiotemporal characteristics of the epidemic; in particular, this deviation would greatly interfere with the analysis of high-incidence areas, which could present great obstacles to the prevention and control of local epidemics.

4. Where/when do spatiotemporal units matter?

4.1. Analysis of points’ distribution pattern

The standard deviation ellipse analysis was carried out on the point of the communities/villages and infected cases in two durations (Fig. 3). From the directional analysis of epidemic transmission, the two durations of the epidemic changed to a certain extent under the three report timestamps, and the directional change of point was relatively weak in the decay duration, which was least obvious in the “symptom onset”. From the clustered characteristics of the epidemic, the infected cases and the communities/villages with the “confirmed” and “quarantined” timestamps in the decay duration were more dispersed. When using the cluster for analysis, the scope and direction of the epidemic in the spread duration are greatly overestimated compared with the analysis of infection cases in the same duration, which might be closely related to the strong tension in the local source of the epidemic.

Based on the results of different types of spatiotemporal units, Ripley’s K function was used to calculate the multiscale distribution characteristics of the local epidemic, and the variation trends of the calculated DiffK with different spatiotemporal units were visualized in the heatmap (Fig. 4). Regardless of which report timestamps or spatial scales were used, the clustered range and intensity of epidemic distribution were higher in the decay duration than in the duration before it. Although the clustered range was relatively small, the clustered intensity was relatively high when the infected cases were counted. However, compared with the cluster layout, which also showed a significant clustered distribution pattern in different report timestamps and durations, when areas were selected as spatial scales, the degree of clustering was no longer significant. It is worth noting in the “symptom onset” statistics, although the number of cases in the spread duration and the decay duration was relatively close, there were still large differences in the clustered range and the clustered intensity, which could confirm the analysis of the above descriptive statistical results.

4.2. Modelling result of the factors

4.2.1. Cross-sectional model recognition and the estimation result

Based on the significant results of the spatial autocorrelation of the model-dependent variables (Z-2.58) and the (robust) Lagrange multiplier test, spatial lag zero-inflated negative binomial regression (SL-ZINB) was further constructed based on zero-inflated negative binomial regression (ZINB), and the Akaike Information Criterion (AIC) value of the model indicates that this improvement could further optimize the most parts of model goodness (LeSage & Pace, 2008; Middela & Ramadurai, 2020). Therefore, in this paper, SL-ZINB (SL-NB in Dalian and Tonghua with area scales due to the lack of excess zero) was selected to estimate and discuss the distance decay effect, the effect of spatial interaction, and the gathering place factor during the COVID-19 epidemic.

Table 2 reports the results of estimating the overall impact factors of the epidemic by using SL-ZINB models. Regardless of which spatial scale was chosen, the source-case distance was always one of the most important and significant factor affecting the local clustering of COVID-19. At the same time, the spatial spillover effect of the dependent variables also significantly promoted the development of the epidemic to a considerable extent. However, compared with the model taking the cluster as the spatial scale, the factors of multiple gathering places had a more significant impact on the estimated results of the model taking the infected cases with neighbourhood-level polygon (area). Catering places and public service facilities played roles in promoting the development of the epidemic with 99% confidence, and this significant effect was reflected only in the “source-case” distance in the model taking the cluster as the spatial scale. Therefore, the “source-case” distance was almost the only factor that significantly affected the transmission of
COVID-19 in the model estimation of the epidemic. Therefore, we could obtain from the overall model estimation results that the expansion of the epidemic in the spatial dimension was mainly related to the "source-case" distance and the gathering of public service places, while the spread of the epidemic was pulled by more kinds of gathering places (Zhang et al., 2021)). There was two interesting point worthy of attention: the first one is when the number of infected cases was taken as the spatial scale, the catering industry had a significant positive effect on the development of the epidemic within two cities, which is possible that individual-level data can better capture the influence of activities with smaller scope and shorter duration, such as dining, on the spread of the epidemic; Secondly, the decisive role of "source-case" distance in the development of epidemic does not seem to exist in Tonghua, which may be caused by the small scale of Tonghua city and the limited potential of distance effect.

### 4.2.2. Longitudinal variation of factors' coefficient

In COVID-19, the distance decay effects showed a trend of increasing at first and then decreasing or stabilizing fluctuations (Figs. 5–7). There was no obvious morphological difference in the time series variation curves of the distance decay coefficient under different spatial scales, but the time-series variation curves of the distance decay coefficient obtained by using three report timestamps had strong heterogeneity. It is worth mentioning that, due to strong manual intervention, the sharp increase of the distance decay coefficient of "confirmed" and "quarantined" during spread duration could be foreseen, but the nonhuman intervention phenomenon of "symptom onset", which conformed to the standard Gaussian distribution, also showed the phenomenon that the coefficient rises during spread duration, which showed that compared with the rapid extension and diffusion in common sense, addition and traceability were the main themes of the spread duration of the epidemic instead.

The spatial lag variable of the epidemic played a significant role in promoting the spread of the epidemic. The estimation results of the impact factors of the epidemic under three kinds of report timestamps were similar to a certain extent: regardless of the type of report timestamp, the positive effect of the spatial lag term on the transmission of the epidemic during the decay duration showed an upwards trend. However, in the spread duration of the epidemic, this homogeneity was no longer obvious, and it was replaced by obvious heterogeneity: the spread of the epidemic during the decay duration showed an upwards trend.

For the variables of the places where people gather, the temporal variation curve of four kinds of gathering place variables would not change obviously when different spatial scales were selected (Fig. 8). However, in the coefficient curve of health-care facilities, this heterogeneity became obvious: when three different report timestamps were selected, the correlation coefficients among different spatial scales had no obvious positive significance, and there rarely was significant negative correlation in the "quarantined" timestamps. Especially during the spread duration of the epidemic, the coefficients and the trends of "confirmed" and "quarantined" differed greatly in the direction of the effect, which was relatively well done by the "symptom onset" data set. Combining the temporal variation of five gathering place coefficients and their correlation coefficients under different spatial scales, the probability and magnitude of this heterogeneity was also the smallest when "symptom onset" was used as a report timestamp. At the same time, the timestamps of "symptom onset" also shows different characteristics from the timestamps of "confirmed" and "quarantined" temporal difference of coefficient estimation of each variable (Fig. 9). At the initial stage of epidemic, the difference between its coefficient and the "confirmed" coefficient is more prominent than the "quarantined" timestamp, which makes it necessary to explore the "symptom onset" timestamp when researching the characteristics and influencing factors of epidemic at the initial phase. It was undoubtedly very important for
the prevention and control of COVID-19 to have an accurate grasp of the impact factors, especially in the initial stage of epidemic spread. Because of these results, researchers and managers should give more attention to the selection of spatial scales when modelling and analysing at the microscale.

5. Which spatiotemporal units matter?

5.1. Report timestamps

First, using different time attributes of case data in analyses could obtain different results, among which “confirmed” and “quarantine” timestamps had similar characteristics in the various spatiotemporal analyses due to their strong manual intervention. However, the
6

“symptom onset” timestamp, which could better reflect the time and process of infection, was often different from the analysis results of the other two report timestamps (Chen et al., 2020; Huang et al., 2020). Because the epidemic under the “symptom onset” timestamp was more normalized, the heterogeneity of characteristics before and after the turning point of the epidemic was not as significant as with the other two timestamp data. Undoubtedly, when analysing or modelling the entire epidemic, choosing the “symptom onset” timestamp data made the analysis results more stable and robust. This phenomenon occurred in the standard deviational ellipse analysis and the temporal variation of modelling results. However, conventional research was mostly based on the “confirmed” timestamp of case data (Guo, Ni, et al., 2020; Han et al., 2021; Liu, Qin, Xie, & Zhang, 2020). A large part of this is due to different research purposes and the difficulties of the data acquisition of the “symptom onset” data and the uncertainty. Therefore, with the deepening of epidemic-related research, the requirements for data mining are increasing (Zhou et al., 2020). Researchers should consider that the “confirmed” timestamp data are deceptive to some extent, and similar data after manual forced intervention cannot accurately reflect the actual transmission of SARS-CoV-2. It was easy to cause cognitive bias or even make an opposite judgement about the epidemic. At this point, the data with the “symptom onset” timestamp were acting better. Therefore, we suggest using more “symptom onset” data in future epidemiological studies, which not only requires academic circles to reach a more consistent consensus but also requires the government to manage the epidemiological information of cases more scientifically and transparently (Sun, Zhang, Yang, Wan, & Wang, 2020; Xia et al., 2020).

5.2. Spatial scales

When selecting spatial scales of cases, through point pattern analysis, there were completely different analysis results on whether the epidemic was clustered and the clustered range of the epidemic. The clustered characteristic of the epidemic was high, and the clustered range was large. Further, clustered cases were often not significant, and the clustered range was relatively small. In the econometric modelling and the temporal variation of the estimation coefficient, the research results obtained by using the epidemic data of different spatial scales could also not explain this problem very well, and it was easy to conclude that the cluster was not significantly affected by the factors of gathering places. At the same time, the infected cases were significantly affected by many factors of gathering places. The effect of place on the transmission of COVID-19 often appeared opposite on the spatial scales. The result of this difference was not unexpected. The spatial scale with points as the epidemic research unit was chosen for its high accuracy, and it could be foreseen (Kwan, 2018). Even so, most of the previous related studies ignored the bias that might be introduced to the research results during the data preparation process and called the research objects “COVID-19 epidemic” in general. The results calculated with these two spatial scales represent two dimensions that should be considered in spatial epidemiology as “spatial sprawling” and “spatial transmission”. The results of this study proved that the two were different and could not be confused. This differentiation in the selection of spatial scales

Table 2
Cross-sectional coefficients estimation with different spatial scales.

| Variable | (a)Beijing(XFD) | (b)Dalian(KSF) | (c)Tonghua(YQL) |
|----------|----------------|----------------|----------------|
| ln(CP)   | -0.0318        | -0.0112        | -0.2868        |
| ln(RA)   | -0.0441        | -0.0230        | -0.0063        |
| ln(SP)   | -0.0104        | 0.0598*        | -0.7272**      |
| ln(PF)   | 0.0506         | 0.0057         | 1.5036***      |
| ln(HCF)  | 0.0482         | 0.0372         | -0.1491        |
| ln(SCD)  | -1.2669***     | -1.0174***     | -2.0279***     |
| Wij(ln(individual)) | 0.0641** | 0.0253         |                |
| Wij(ln(Area))    | 0.0404         | 0.1902***      |                |
| Constant   | 172.3          | 517.9831       |                |
| Log likelihood| -571.9831     | -36.3          |                |
| AIC       | 1082.32        | 1069.78        |                |
| AIC_C    | 1082.32        | 1069.78        |                |
| AIC_C_C  | 1082.32        | 1069.78        |                |
| AIC_C_C_N| 1082.32        | 1069.78        |                |
| AIC_C_C_N | 1082.32        | 1069.78        |                |

*a p < 0.1 **p < 0.05 ***p < 0.01.

Fig. 4. Trends of DiffK with different spatiotemporal units in the two durations.
reduced the reference significance of these studies in the subsequent reference process and greatly lost the regional comparative value of these studies (Liu, Qin, Xie, & Zhang, 2020; Rex, de Souza Borges, & Kafer, 2020; Wu et al., 2020).

Finally, the impact of the size and shape of epidemic spatial boundaries on the results of spatiotemporal epidemiological studies was also reported in this paper. In the past, related studies were mostly macro- or large-scale studies, their research units were usually administrative regions or large grid units, and the selection principle of data spatial boundaries caused less disturbance in their research results (Arab-Mazar, Sah, Rabaan, Dhama, & Rodriguez-Morales, 2020; Giuliani, Dickson, Espa, & Santi, 2020; Michelozzi et al., 2020; Xiong, Wang, Chen, & Zhu, 2020). In contrast, the above results can indirectly reflect the differences brought about by the selection principles of different spatial boundaries. The spatial modelling results of econometric models provide sufficient evidence for the significant spatial spillover effect of the COVID-19 epidemic at the microlevel. This significant spillover effect will undoubtedly greatly affect the estimation results when the size or shape of the research units changes (LeSage and Pace, 2008). Therefore, we suggest that the smallest areal unit of data or grids with a small area should be used as often as possible, as they are not limited to arbitrary spatiotemporal units when individual-level space-time data are not available. It is necessary to discuss the modifiable areal unit problem (MAUP) and the uncertain geographic context problem (UGCoP) in future microscale spatiotemporal epidemiological studies (Dungan et al., 2002; Kwan, 2012, 2018). Based on this, perhaps academic circles should pay more attention to the role of remote sensing and raster data in epidemiological research because of their high-resolution characteristics.

5.3. Limitations and uncertainties

The epidemic was modelled based on distance, spatial effect, and places where people gather; however, the model did not include all the factors that affect the transmission of epidemics, such as population density, economic activity intensity, and population mobility. Second, although this study proves that the selection of spatiotemporal units was very important in the microlevel study of the COVID-19 epidemic, the results of this study did not provide enough evidence to show whether the clusters or infected cases are better or worse as the spatial scale of the epidemic research. The discussion based on the cluster could better reflect the “sprawling” process of the epidemic in the spatial dimension, while the number of infected cases can better explain the focus and bias of the epidemic, which needs much follow-up for validation. In addition,
it should not be neglected that the open-source case data independently counted by the Municipal Health Commissions and the Centers for Disease Control and Prevention in this study were not the most detailed and accurate. Although we approximated the data using more consistent and scientific principles, the possible errors or bias caused by approximate processing in such microscopic research is still worth discussing, and the accurate location of relevant case points needs to be further corrected in the government’s internal data. Finally, although COVID-19 was undoubtedly a typical space pandemic, the results of this study can only be regarded as enlightening and alarming work. The heterogeneity caused by different spatiotemporal units and their advantages and disadvantages need larger-scale inspection work and additional types of studies.

6. Conclusion

Our research highlights the importance of the selection of spatiotemporal units in microgeographic epidemiology studies through the multidimensional spatiotemporal analysis of COVID-19 in three typical localized outbreaks in mainland China. Different report timestamps and spatial scales had a significant impact on the results of spatiotemporal analysis and modelling. This heterogeneity was reflected in the clustering and directionality of epidemic spatial distribution, its factors, and their temporal variation. According to the results of quantitative studies, due to the better robustness of the timestamp “symptom onset”, we suggest that more attention should be given to it instead of focusing only on “confirmed” timestamps because “confirmed” data are often deceptive and unstable in epidemiological research. Spatial scales were also worth emphasizing in the microstudy of the epidemic in the individual units. Based on this, the specific content of the follow-up epidemiological study should be clarified more clearly; that is, the spatial transmission should not be confused with the spatial “sprawling” of epidemics, which will greatly reduce the possibility of systematization of research results. In addition, the modifiable areal unit problem (MAUP) and the uncertain geographic context problem (UGCoP) are also worthy of a more in-depth discussion in the microscopic study of epidemics.

Authors’ contributions

Sui Zhang: Conceptualization, Data curation, Methodology, Data analysis, Visualization, Software, Writing - original Draft, Writing - review & editing. Minghao Wang: Data curation. Zhao Yang: Data curation. Baolei Zhang: Supervision, Project administration, Funding acquisition, Writing - review & editing.
Fig. 7. Temporal variation of the coefficient of factors in the Tonghua (YQL) outbreak.

Fig. 8. Pearson’s $r$ between the temporal trends of coefficients with different spatiotemporal units: (a) Beijing (XFD), (b) Dalian (KSF), (c) Tonghua (YQL).
Fig. 9. Temporal characteristics of standardized differentiation between coefficients with different report timestamps: (a)Beijing (XFD), (b)Dalian (KSF), (c)Tonghua (YQL).

Declaration of competing interest

The authors declare that they have no conflict of interest.

Acknowledgments

This research was funded by the National Social Science Foundation of China (No. 18BJY086). The authors wish to gratefully acknowledge the anonymous reviewers who helped to improve this paper through their thorough review.

References

Arab-Mazar, Z., Sah, R., Rabaa, A. A., Dhana, K., & Rodriguez-Morales, A. J. (2020). Mapping the incidence of the COVID-19 hotspot in Iran – implications for travellers. Travel Medicine and Infectious Disease, 34, 101630.

Bertazzi, A. L., Franco, E., Mohler, G., Short, M. B., & Sledge, D. (2020). The challenges of modelling and forecasting the spread of COVID-19. Proceedings of the National Academy of Sciences of the United States of America, 117(29), 16732-16738.

Bertuzzo, E., Mari, L., Pasetto, D., Miccoli, S., Casagrandi, R., Gatto, M., et al. (2020). The geography of COVID-19 spread in Italy and implications for the relaxation of confinement measures. Nature Communications, 11(1).

Birch, C. P. D., Oom, S. P., & Beecham, J. A. (2007). Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. Ecological Modelling, 206 (3–4), 347-359.

Boots, B. N., & Getis, A. (1988). Point pattern analysis. Morgantown: Regional Research Institute, West Virginia University.

Brito-Redon, A., & Serrano-Aroca, A. (2020). A spatio-temporal analysis for exploring the effect of temperature on COVID-19 early evolution in Spain (p. 728). Science of the Total Environment.

Burger, M., Van Oort, F., & Linders, G.-J. (2009). On the specification of the gravity model of trade: Zeros, excess zeros and zero-inflated estimation. Spatial Economic Analysis, 4(2), 167-190.

Chan, J. F.-W., Yuan, S., Kok, K.-H., To, K. K.-W., 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.

Chen, Y., Liang, W., Yang, S., Wu, N., Gao, H., Sheng, J., et al. (2021). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet, 395(10223), 497-506.

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., et al. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. Lancet, 395(10223), 507-513.

Chun, T. W., Carnuth, L., Finzi, D., Shen, X., DiGiuseppe, J. A., Taylor, H., et al. (1997). Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. Nature, 387(6629), 183-188.

Dungan, J. L., Perry, J. N., Dale, M. R. T., Legendre, P., Citron-Pousty, S., Fortin, M. J., et al. (2002). A balanced view of scale in spatial statistical analysis. Ecography, 25(5), 626-640.

Flowerdew, R., & Aitkin, M. (1982). A method of fitting the gravity model based on the Poisson distribution. Journal of Regional Science, 22(2), 191-202.

Flowerdew, R., & Boyle, P. J. (1995). Migration models incorporating interdependence of movers. Environment & Planning A, 27(9), 493-502, 1.

Franch-Pardo, I., Napoleton, B. M., Ronet-Verge, F., & Billa, L. (2020). Spatial analysis and GIS in the study of COVID-19. A review (p. 739). Science of the Total Environment, 626(6), 640.

Gao, S., Rao, J. M., Kang, Y. H., Liang, Y. L., Kruse, J., Dopfer, D., et al. (2020). Association of mobile phone location data indications of travel and stay-at-home mandates with COVID-19 infection rates in the US. JAMA Network Open, 3(9).

Garrett, A. C., & Bailey, T. C. (1996). Interactive spatial data analysis in medical geography. Social Science & Medicine, 42(6), 843-855, 1982.

Getis, A. (1984). Interaction modeling using second-order analysis. Environment & Planning A: Economy and Space, 16(2), 173-183.

Giuliani, D., Dickson, M. M., Espa, G., & Santì, F. (2020). Modelling and predicting the spatio-temporal spread of coronavirus disease 2019 (COVID-19) in Italy. SSRN Electronic Journal.

Grushka, E. (1972). Characterization of exponentially modified Gaussian peaks in chromatography. Analytical Chemistry, 44(11), 1733-1738.

Guang, W., Ni, Z., Hu, Y., Liang, W., Ou, C., He, J., et al. (2020). Clinical characteristics of coronavirus disease 2019 in China. New England Journal of Medicine, 382(18), 1708-1720.

Guo, D., Wang, D., Hallegatte, S., Davis, J. S., Hao, J., Li, S., et al. (2020). Global supply-chain effects of COVID-19 control measures. Nat Hum Behav.

Han, Y., Yang, L., Jia, K., Li, J., Feng, S., Chen, W., et al. (2021). Spatial distribution characteristics of the COVID-19 pandemic in Beijing and its relationship with environmental factors (p. 761). Science of the Total Environment, 633, 1-10.

He, F., Deng, Y., & Li, W. (2020). Coronavirus disease 2019: What we know? Journal of Medical Virology, 92(7), 719-725.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet, 395(10223), 497-506.

Kraemer, M. U. G., Yang, C.-H., Gutierrez, R., Wu, C.-H., Klein, B., Pigott, D. M., et al. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. Science, 368(6490), 493–5.

Kwan, M.-P. (2012). The uncertain geographic context problem. Annals of the Association of American Geographers, 102(5), 956-968.

Kwan, M.-P. (2018). The limits of the neighborhood effect: Contextual uncertainties in geographic, environmental health, and social science research. Annals of the Association of American Geographers, 108(6), 1482-1490.

Lee, D., Robertson, C., & Marques, D. (2021). Quantifying the small-area spatio-temporal characteristics of the COVID-19 pandemic in Scotland during a period with limited testing capacity. Spatial statistics, 100508, 100508.

Leeve, D. W. (1926). Measuring geographic concentration by means of the standard deviational ellipse. American Journal of Sociology, 31(1), 88-94.
