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.
Measurement of contagion spatial spread probability in public places: A case study on COVID-19

Lu Chen a, Xiuyan Liu a,b, Tao Hu c, Shuming Bao d,e, Xinyue Ye f, Ning Ma g, Xiaoxue Zhou h,*

a School of Management and Economics, Southeast University, Nanjing, 211189, China
b National School of Development and Policy, Southeast University, Nanjing, 211189, China
c Department of Geography, Oklahoma State University, Stillwater, OK, 74075, USA
d Ecomputation Center for Social Science, Wuhan University, Hubei, 430079, China
e Center for Geographic Analysis, Harvard University, Cambridge, MA, 02138, USA
f Geo-computation Center for Social Science, Wuhan University, Hubei, 430079, China
g School of Law, Southeast University, Nanjing, 211189, China
h Jiangsu Institute of Industrial Development, Nanjing University of Finance and Economics, Nanjing, 210023, China

ARTICLE INFO
Keywords: Spread probability Covid-19 Public places Spatial analysis Wasserstein distance

ABSTRACT
The scale and scope of the COVID-19 epidemic have highlighted the need for timely control of viral transmission. This paper proposed a new spatial probability model of epidemic infection using an improved Wasserstein distance algorithm and Monte Carlo simulation. This method identifies the public places in which COVID-19 spreads and grows easily. The Wasserstein Distance algorithm is used to calculate the distribution similarity between COVID-19 cases and the public places. Further, we used hypothesis tests and Monte Carlo simulation to estimate the spatial spread probability of COVID-19 in different public places. We used Snow’s data to test the stability and accuracy of this measurement. This verification proved that our method is reliable and robust. We applied our method to the detailed geographic data of COVID-19 cases and public places in Wuhan. We found that, rather than financial service institutions and markets, public buildings such as restaurants and hospitals in Wuhan are 95 percent more likely to be the public places of COVID-19 spread.

1. Introduction
Since January 2020, the SARS-CoV-2 Disease (COVID-19) has significantly impacted individuals’ health and the socioeconomic vitality of the whole world. As of March 1st, 2022, there were over 438 million confirmed cases and 5.98 million total deaths globally. Person-to-person interaction through respiratory droplets is the main channel for the spread of COVID-19. It is critical to understand the potential transmission dynamics of COVID-19 within the building environment ecosystem and the human behavior, spatial dynamics, and building operational factors that potentially promote and mitigate the spread and transmission of COVID-19 (Dietz et al., 2020). In consideration of this, researchers have focused on the impact of spatial variation characteristics on the spread of COVID-19, including urban density, population density, ethnicity and interactive mode (Megahed & Ghoneim, 2020; Elson et al., 2021; Li et al., 2021; Liu & Yang, 2022; Siljander et al., 2022; Sun & Zhai, 2020) and built environment (Fang et al., 2020; Nguyen et al., 2020; Rocklov & Sjödin, 2020; Kashem et al., 2016).

Inside cities, the built environment factors influence the spread of COVID-19. Such factors include vehicle road and bicycle lanes, sidewalks, and public green space (Capolongo et al., 2020; Hishan et al., 2020). For example, Kimmelman (2020) argued for increased bicycle and active transportation infrastructure to be designed for multimodal options. In the study of dilapidated buildings, sidewalks, and green landscaping through Google Street View, Nguyen et al. (2020) found environment of single-lane residential roads and green streets tended to have fewer COVID-19 cases, while multi-family home areas with physical disorders had more COVID-19 cases. Others made the case for less household crowding (Emeruwa et al., 2020; Hu et al., 2021).

Beyond the above factors, consistent with the mechanisms of spread, public spaces with the characteristic of individuals’ “stays” has more impacts on the spread of COVID-19, such as hospitals, restaurants, banks, and supermarkets. Individuals travel to such places routinely and stay for a period of time. Undoubtedly, such trips will increase the...
probability of person-to-person contact and infection. From this point of view, it is of great significance to evaluate the impacts of such places on the spread of COVID-19 in order to identify the category of public space with the highest probability of being infected. Goldfarb and Tucker (2020) examined the patterns of consumers visiting retail stores and what kinds of retailers present the most risk. They find that small businesses, particularly single store retailers, attract fewer visits and should therefore open first. Benzell et al. (2020) measured the relative transmission reduction benefit and social cost of closing 26 categories of locations including shops, entertainment, and service providers. Atkinson et al. (2020) developed a Mobility and Engagement Index to measure the impact of COVID-19 on economic activity.

Some studies on the spatial distribution of the epidemic also try to explore the relationship between the epidemic and different public places, and the mainstream method used is the geographically weighted regression (GWR). Li et al. (2020) used the GWR model to study how different urban factors influence the distribution of COVID-19 in Wuhan. Hamidi et al. (2020) applied GWR to explore the relationship between COVID-19 cases and seven different public places in Beijing. Based on these studies, we attempt to answer the simple question of which public places are most likely to result in the spread of COVID-19. We utilized a new spatial probability algorithm method of epidemic infection, employing improved Wasserstein distance algorithm and Monte Carlo simulation. More broadly, we used the crawled geographic data of COVID-19 cases to measure the similarity between the spatial distribution of cases and that of public places. By doing so, we seek to help ascertain which public places may be the main spatial spreading media for the epidemic. Our work provides suggestive data and an empirical framework for helping the government decide which economic activity should reopen first and which should remain shut down until later time.

The viral infection rates vary dramatically across areas of different densities. For instance, Hamidi et al. (2020) found that larger metropolitan areas of high density have higher COVID-19 infection and mortality rates than county areas. Other evidence suggests that the spread of COVID-19 cases is not always in line with density. Dense cities and countries like Singapore, Hong Kong, and South Korea have fewer case rates than the United States (Beech & Anseele, 2020; Fang & Wahba, 2020).

2. Data

We collected the residence information of the COVID-19 cases, as well as the geographic coordinates of all public facilities in Wuhan, China.

Because of privacy concerns, in most countries, only official health authorities have access to residential information on patients. The Chinese city of Wuhan was sealed off on January 23rd, 2020, restricting the mobility of humans in the city. Therefore, we assume that the confirmed cases reported after that date should have spread infection through communities before the lockdown.

Since early February 2020, the city of Wuhan has reported the confirmed and suspected infection data of residents in each gated community through daily bulletins. Based on such information, this paper collected the case data of Wuhan city from February 3rd to February 27th and acquired the residential address of each case through the decoding of AutoNavi (https://lbs.amap.com). We collected a panel of 13833 COVID-19 cases, which were distributed in the 456 gated communities.

We also collected more than 250,000 commercial point-of-interest (POIs) of Wuhan city from the Gaode Map, which covers hospitals, financial services, schools, shopping malls, restaurants, and entertainment sites.

3. Methodology

The core idea of this study is to test the statistical significance of the similarity between the spatial distribution of epidemic cases and public places. We define this statistical significance as the probability of transmission in certain public places: the Spread Possibility (SP) Index. Fig. 1 illustrates this work flow.

1. Spatial similarity estimation based on improved Wasserstein distance algorithm

Wasserstein Distance, also known as Optimal Transport Distance or Earth Mover’s Distance, is a distance measure defined in a given metric space $M$ that measures the similarity between two probability distributions. The idea of Wasserstein Distance is classical. Wasserstein Distance is widely used in probability theory and mathematical statistics in the field of mathematics. In addition to mathematicians’ theoretical importance, researchers have provided a successful framework for the comparison of objects in fields of application (Panaretos & Zemel, 2019), such as pharmaceutical statistics (Munk & Czado, 1998), image retrieval (Rubner et al., 2000), computer vision (Ni et al., 2009), genomics (Evans & Masen, 2012), and economics (Billings & Johnson, 2016). Due to its excellent mathematical characteristics, it has been introduced into artificial intelligence fields such as facial recognition and picture analysis in recent years (Carlsson et al., 2018; Martin et al., 2017). In general, this method has been widely used in many fields, but has not been effectively applied in geography.

Wasserstein distance is an abstraction of two specific problems: the optimal transportation distance and the bulldozer distance. For probability distributions $\mu(x)$ and $\nu(y)$ defined on $\mathbb{R}^n$, which are the marginal distributions of the joint distribution $\Gamma(\mu, \nu)$ defined on $\mathbb{R}^n \times \mathbb{R}^n$. For $p \geq 1$, the Wasserstein distance between $\mu(x)$ and $\nu(y)$ can be defined as:

$$ W_p(\mu(x), \nu(y)) := \left( \int_{\mathbb{R}^n \times \mathbb{R}^n} d(x,y)^p \gamma(x,y) \, dx \, dy \right)^{1/p} \tag{1} $$

The original calculation method for solving Wasserstein distance is a linear programming approach, where the key objective is to find the minimum distance scheme in the joint distribution $\Gamma(\mu, \nu)$. $\gamma(x,y)$ is an element of $\Gamma(\mu, \nu)$. Similar to the method of Billings and Johnson (2016) for calculating the distribution distance between two discrete industries, we normalize the distributions before the calculation.

Drawing on the method of Cuturi (2013), which smooths the classic optimal transport distance computed through Sinkhorn’s matrix scaling algorithm with an entropic regularization term, and further modified by Carlsson et al. (2018), the Wasserstein distance with an entropic regularization is:

$$ W_p(\mu(x), \nu(y)) := \left( \int_{\mathbb{R}^n \times \mathbb{R}^n} d(x,y)^p \gamma(x,y) \, dx \, dy - \gamma(\gamma) \right)^{1/p} \tag{2} $$

where $\gamma(\gamma)$ is the entropic regularization of the joint distribution $\Gamma(\mu, \nu)$, $\alpha$ is entropy constraint coefficient and $\alpha > 0$ (Léonard, 2012; Solomon et al., 2015).

$$ \gamma(\gamma) = -\int_{\mathbb{R}^n \times \mathbb{R}^n} \gamma(x,y) \ln \gamma(x,y) \, dx \, dy \tag{3} $$

By entropic regularization, function (2) is a strictly convex function with a unique optimal solution.

Further, the kernel density function is defined as:

$$ K(x, y) = e^{-\| d(x,y) \|^p} \tag{4} $$

For any $t > 0$, function (4) is positive definite. Substituting equation (4) into equation (2) gives:
Formula (5) is the Wasserstein distance with an entropic regularization term, based on the Sinkhorn algorithm. In the actual computation process used in this paper, we set $t = 1/\alpha$ and $\alpha = 1$ in this paper (Solomon et al., 2015).

Specifically, in this study, $p$ is set to be 2 when studying the spreading probability of an epidemic in two-dimensional space. We compared the epidemic spatial distribution with the distribution of public places associated with a set of spatial observations on $\mathbb{R}^2$, which we denote as $f_j$ and $f_k$, respectively. The Wasserstein distance between these distributions, denoted as $W_{j,k}$, is defined as:

$$W_{j,k} = \left( \inf_{\gamma(x,y) \in \Gamma(f_j, f_k)} \int_{\mathbb{R}^n \times \mathbb{R}^n} d(x,y)^2 \gamma(x,y) dx dy \right)^{1/2}$$  \hspace{1cm} (6)

The entropic regularization is added in equation (6) gives:

$$W_{j,k} = \left( \inf_{\gamma(x,y) \in \Gamma(f_j, f_k)} \int_{\mathbb{R}^n \times \mathbb{R}^n} d(x,y)^2 \gamma(x,y) dx dy \int_{\mathbb{R}^n \times \mathbb{R}^n} \gamma(x,y) \alpha \ln(\gamma(x,y)) - \ln(\kappa(x,y))/t \right)^{1/2}$$  \hspace{1cm} (7)

4. Spread probability index based on Monte Carlo simulation

To examine whether the epidemic is likely to spread in space through such public places $k$, a spread probability index between the COVID-19 cases (denoted as $j$) and the public places $k$ needs to be conducted based on counterfactual samples.

Our **Null Hypothesis** is that no spatial similarity occurs between COVID-19 cases $j$ and the public places $k$ needs to be conducted based on counterfactual samples.

Consistent with Billings and Johnson (2016), we construct our counterfactual of randomly located (pseudo) public constructions based on two specific criteria: 1) the sample should be drawn from the set of locations where a public place could potentially locate, and 2) the
sample size used in constructing the counterfactual must be equal to the number of the public place category. This strategy helps control undevelopable land as well as other unobservable constraints on public place locations. In selecting the set of locations for our counterfactual, we restrict pseudo-public construction to the sites of our full population of establishments. For each ordered pair \((j, k)\), we construct pseudo-public constructions distribution \(k\) based on a random selection of \(N_k\) locations from the set of all establishments.

Therefore, our counterfactual measure of spatial similarity \(W_{j,k}\) incorporates the spatial distribution of COVID-19 cases compared to pseudo-public constructions \(k\) of size \(N_k\). The computation \(W_{j,k}\) is repeated for 1000 pseudo-public place categories to generate an empirical null distribution for the correlation of patients to pseudo-public places. Set \(A\) is defined as the set of 1000 simulation results \(W_{j,k}\). We define the Spread Possibility Index (SP index) as follows:

\[
B = \{ W_{j,k} | W_{j,k} > W_{j,k} \}_{W_{j,k} \in A}
\]

\[
SP = \frac{\text{Card}(B)}{\text{Card}(A)}
\] (8)

In formula (8), we construct our SP Index using the occurrence of \(W_{j,k}\) less than \(W_{j,k}\). For instance, if \(W_{j,k}\) is less than \(W_{j,k}\) 950 times, then we assign the SP Index as 0.95. Therefore, the magnitude of the SP Index represents the statistical significance (greater than or equal to 0.95) of spatial similarity for COVID-19 spatial distribution to public place category \(k\).

Compared with the GWR model (Han et al., 2021; Li et al., 2020), our method is a nonparametric estimation. Our method directly examines how similar the two spatial distributions are by using the statistical methods of hypothesis testing and Monte Carlo simulations. The GWR method is a regression method based on the linear relationship between variables, which assumes that the stochastic disturbance team follows a normal distribution. Unlike this approach, our model does not require this prior assumption, nor does it need to set up a linear model first. But the two approaches can complement and verify each other for the same problem.

5. Validation: Snow data set

Spatial analysis of epidemics was first used by John Snow in his research about the outbreak and spread of cholera in Soho, London in the mid-1800s (Snow, 1855). He mapped the distribution of cases and the location of pumps, visually revealing the relationship between contaminated well water and cholera outbreaks, hence, overturning the previous miasma statement. From then on, such spatial analysis method has been widely used in studying the source and spread spaces of infectious epidemics. We used data from Snow’s study to test the reliability and robustness of our measurement.

Referring to the practice of Pare et al.’s study (Pare et al., 2020), we use Snow’s classical and real data to verify the accuracy and robustness of the model proposed in this paper. The GIS data for Snow’s work is made public by Caitlin Dempsey in 2013 (https://www.gislounge.com/john-snows-cholera-map-gis-data/). In Fig. 2, the map is shown, where each green diamond mark corresponds to death at that address, and the red octagonal marks illustrate the locations of the pumps.

Firstly, we calculated the Wasserstein distance between the spatial distribution of cholera cases and actual pumps in snow data, which is 0.0277 based on the data set. Secondly, we randomly generated 8 pumps in the range \((18.0450° E – 18.1300° E, 52.8950° N – 52.9950° N)\) in the figure (black dotted box in Fig. 2). In order to verify the conclusion of this method is robust and reliable.

In order to test the robustness of the calculation results, we simulated this calculation process repeatedly for 100 times, the density distribution of the results for 100 times is shown in Fig. 3. We can see that the 100 simulation results are mainly distributed between 98% and 99%. It is proved that the conclusion of this method is robust and reliable.

6. Results

Fig. 4 depicts the spatial distribution of COVID-19 cases in Wuhan on February 10th and 25th, 2020. We used the Kernel Density method\(^2\) to calculate the probability of the pumps in Soho district being the spatial transmission medium of cholera was about 98%. Using the 95% confidence level, we reject the null hypothesis that there is no spatial distribution similarity between cholera cases and pumps, so it can be inferred that pumps are an important medium for cholera transmission in this region. Our research conclusion is consistent with that of Pare et al. (2020) using discrete-time SIS model to analyze Snow data. This means that our measurement method is effective in the application of Snow data sets.

\(^1\) We must admit that this sampling method has certain errors, because in this space there are a lot of places are not suitable for construction of pumps. We should put the position of all public places in the region as a sampling data set to calculate the accurate results. This sampling strategy is used in the following COVID-19 transmission study.

\(^2\) The Kernel Density analysis is implemented by ArcGIS Pro based on the quartic kernel function with the bandwidth of 0.02789 mile.
plot the spatial thermal map distribution of the two days. Jianghan district, Qingshan District, Wuchang District, and Hongshan District were the main areas of the urban epidemic outbreak, and gradually spread from the above four areas to most of the urban areas in half a month.

In this article, we focused on the transmission role of hospitals, banks, restaurants, and other public places in the epidemic (Fig. 5). One characteristic of such public places, which is just like the pumps in Snow’s study (1855), is that people are more likely to go to such places that are closer to them. This shows that once such places become
transmission media, the spatial distribution of such places will be very similar to that of the epidemic. It should be noted that there may be a large bias using the method depending on observation and conjecture in Snow’s research. To improve Snow’s spatial analysis method, we used the specific number of infected people in each gated community as a weight in actual calculations and use a statistical inference method to strictly demonstrate the similarity between the distribution of different public places and cases, which makes the conclusion more accurate.

Our basic results are reported in Table 1. We respectively calculate the cumulative distribution of cases on February 10th and February 25th. It can be found that the results measured at the two timings are very robust. In Table 1, we report the number of public buildings of 15 categories in a given space, the Wasserstein Distance of the spatial distribution of cases, and the SP Index calculated using Monte Carlo simulation.

As shown in Table 1, the probability of the COVID-19 epidemic spreading in the three sub-categories of restaurants is 1, which means that the spatial distribution of Chinese Restaurants, Snack Bars, and Foreign Restaurants is significantly similar to that of cases at a significance level of more than 99%. Such results indicate that restaurants are major places of infection of the epidemic.

Among the three subcategories of Medical Centers, the SP index of General Hospitals and Specialized Hospitals is equal to 1, and the SP index of Clinics is 0. From the number of these sub-categories, we can judge that category number cannot affect the similarity of spatial distribution. It can be judged that the spatial distribution is the only factor influencing the similarity of the spatial distribution of public places and cases.

For the subcategories of financial services, the SP Index values of Insurance Services, Banks, and Post Offices are 0.055, 0.263, and 0.006, respectively. Using a standard significance level (95%), we cannot reject the null hypothesis that each spatial distribution of these three kinds of public places is not similar to that of COVID-19 cases. Therefore, there is a high probability that these three sub-categories of public places are not the medium of COVID-19 transmission.

For the public facility of Education, we mainly examine two subcategories, schools and training organizations. We got two diametrically opposite results: Schools’ SP Index was 0, and Training Organizations’ SP Index was 1.

In the general category of Markets, the results of the SP Index of supermarkets and convenience stores are close to 0, indicating that markets did not act as a spread medium. In contrast, cinemas and stadiums in the Entertainments classification have a statistically significant high probability of being the spreading medium of the epidemic.

Table 1

| Category       | Sub Category       | Number | February 10th | February 25th |
|----------------|--------------------|--------|---------------|---------------|
|                |                    |        | SP Index      | Wasserstein Distance |
| Restaurants    | Chinese Restaurants| 35482  | 1             | 0.1392        |
|                | Snack Bars         | 4999   | 1             | 0.1236        |
|                | Foreign Restaurants| 1804   | 1             | 0.1304        |
| Medical Centers| Clinics            | 1906   | 0             | 0.1434        |
|                | General Hospitals  | 1317   | 1             | 0.1325        |
|                | Specialized Hospitals| 945  | 1             | 0.1316        |
| Financial Services | Insurance Services | 1437  | 0.055         | 0.1415        |
|                | Banks              | 1627   | 0.263         | 0.1406        |
|                | Post Offices       | 270    | 0.006         | 0.1456        |
| Education      | Schools            | 6194   | 0             | 0.1419        |
|                | Training Organizations | 5624 | 1              | 0.1361        |
| Markets        | Super Markets      | 1713   | 0.001         | 0.1427        |
|                | Convenience Stores | 11277  | 0             | 0.1453        |
| Entertainments  | Cinemas            | 247    | 1             | 0.1289        |
|                | Stadiums           | 948    | 1             | 0.1339        |

7. Discussion

In the studies of Goldfarb and Tucker (2020), Atkinson et al. (2020) and Benzell et al. (2020), the frequency of visits and the length of stay in public places such as supermarkets have become the main factors for assessing the risk of COVID-19 spreading in a given place. However, there is a priori hypothesis for this criterion: the interaction behavior of all people in public places is undifferentiated. Nonetheless, there are obvious differences in interactions between populations in different public places. It is noteworthy that the criteria for determining hazardous public places used in this paper are derived from the results of community transmission, which is fundamentally different from the previous studies.

Consistent with many studies and news about the spread of COVID-19 (Han et al., 2021; Li et al., 2020), our finding discovers that restaurants present a high probability of being the spreading space of the epidemic. In such spaces, consumers usually stay for a long time and communicate in a relatively small and closed space. This scenario can explain the spreading of the epidemic in public places such as entertainment centers.

For medical centers, general hospitals and specialized hospitals have a higher probability of being the spreading space of the epidemic, while clinics present a low probability. Our conclusions are basically consistent with the conclusions of Li et al. (2020) and Han et al. (2021). The difference is that we discussed hospitals and clinics separately. One possible reason for the results is that hospitals are the main treatment agencies for infected cases, greatly increasing the risk of such places. As can be seen from Fig. 6, the spatial distribution of hospitals and cases is indeed highly similar, and areas with high hospital density also have high morbidity. Besides, since China has relatively imperfect hierarchical diagnosis and treatment systems, many patients directly go to the general hospitals and specialized hospitals for treatment, reducing the risk of being infected in clinics. Definitely, in other countries, any type of medical center is a risky place, and medical staff must protect themselves well. Not only do hospital facilities have limited options for social-distancing measures to prevent infectious spread, but health care facilities also often house patients with vastly different requirements from the environment around them (Dietz et al., 2020).

Spatial configuration of buildings can encourage or discourage social interactions. In public places where financial services are provided, such places are generally more spacious, and there are glass partitions...
between staff and customers, so there is less communication between customers. Besides, the spaces of such places are generally larger with clean and tidy environments, making these places less likely to be the spreading space of the epidemic. Intuitively, the similarity between the spatial distribution of banks and the distribution of COVID-19 cases is less than that of hospitals (Fig. 7). Therefore, in our calculation results, the places mentioned above have a low probability of transmission as a new crown epidemic. This is different from the conclusions of Han et al.
(2021), and in some practical cases, there are indeed few news reports of patients infecting themselves in banks.

As for places of education, the spreading probability is relatively low in schools since students are on winter vacation, while the training organizations present more risk because they are the main places for Chinese students to study and interact on vacation. Due to the time interval of the data, the conclusions of our study in educational settings are different from those of Han et al. (2021). In fact, schools are very easy to turn into places for the spread of the epidemic, which has been confirmed by news reports from China, the United States, and the United Kingdom.

The results seem unconvincing since the SP index of supermarkets and convenience stores are close to zero, which is unintuitive. The first large-scale infection in Wuhan was found in a seafood supermarket, making it easy for us to accept markets as a high-risk place. However, the frequency of social interaction and probability of long-time stay in supermarkets is very low while individual mobility is high making such places less likely to be the spreading space of COVID-19. In fact, except for the initial cases of infection in a seafood market in Wuhan, there has not been any news of similar large-scale cluster infections occurring in other supermarkets and markets, which is inconsistent with the small-scale outbreak in Beijing in June 2020. In the process of community spread, the probability of Markets being the spreading space of COVID-19 is not too high.

8. Conclusions

This paper designed an estimation method based on the spatial distribution of different public places being the spreading space for COVID-19. This method mainly adopted the Wasserstein Distance algorithm and developed a specific spreading probability index, SP Index, via statistical hypothesis testing, and Monte Carlo simulation.

We use the data of Wuhan city to demonstrate the paradigm workflow of methodology, calculation, and result analysis. It should be admitted that the data we used is not comprehensive, considering the protection of patients’ privacy and the lag in news coverage. However, the conclusions drawn from this paper must be noted by policymakers for epidemic prevention and control. If our method can be applied to comprehensive and complete epidemic data, the conclusions of such spatial analyses will be very instructive for epidemic control and minimizing the damage thereof to the economy and convenience of individuals.

Author statement

The corresponding author is responsible for ensuring that the descriptions are accurate and agreed by all authors. Lu Chen: Conceptualization, Methodology, Software. Xiuyan Liu: Conceptualization, Methodology, Software. Tao Hu: Data curation, Writing- Original draft preparation. Shuming Bao: Supervision. Xinyue Ye: Supervision. Ning Ma: Writing- Reviewing and Editing. Xiaoxue Zhou: Writing- Reviewing and Editing.

Acknowledgements

The authors would like to acknowledge Harvard Dataverse from Harvard University and China Data Lab (http://chinadatalab.org/) for providing access to the COVID-19 data. Our research is funded by Soft Science Research Program of Zhejiang Province (NO.2020C3S5SA01366). We would like to thank Jialei Wang from Wuhan University and Yuan Jun from NJIT for their help in writing this paper. The authors wish to gratefully acknowledge the anonymous reviewers who helped to improve this paper through their thorough review.

References

Atkinson, T., Dolman, J., Koch, C., Koening, E., Mertens, K., Murphy, A., & Yi, K. (2020). Mobility and engagement following the SARS-CoV-2 outbreak. Federal Reserve Bank of Dallas. Working Papers 2020.

Beech, N., & Anseele, F. (2020). COVID-19 and its impact on management research and education: Threats, opportunities and a manifesto. British Journal of Management, 31 (3), 434–449.

Benzell, S. G., Collis, A., & Nicolaides, C. (2020). Rationing social contact during the COVID-19 pandemic: Transmission risk and social benefits of US locations. Proceedings of the National Academy of Sciences, 117, 14642–14644.

Billings, S. B., & Johnson, E. B. (2016). Agglomeration within an urban area. Urban Economics, 91, 13–25.

Capolongo, S. S., Rebecchi, A. A., Buffolli, M. M., Appolloni, L. L., Signorelli, C. C., Fara, G. M. G. M., & D Alessandro, D. D. (2020). COVID-19 and cities: From urban health strategies to the pandemic challenge. A Decalogue of Public Health Opportunities. Acta bio-medica de L’Ateneo Parmense, 91, 13–22.

Carlsson, J. G., Behrozii, M., & Mihic, K. (2018). Wasserstein distance and the distributionally robust tsp. Operations Research, 66, 1603–1624.

Dietz, L., Horve, P. F., Coll, D. A., Fretz, M., & Wymelenberg, K. (2020). 2019 novel coronavirus (Covid-19) pandemic: Built environment considerations to reduce transmission. mSystems, 5(2), e00245-20.

Elsn, R., Davies, T. M., Lake, I. R., Vivencos, R., Blomquist, P. B., Charlett, A., & Dahvrea, G. (2021). The spatio-temporal distribution of COVID-19 infection in England between January and June 2020. Epidemiology and Infection, 149, e73.

Emerwuwa, U. N., Ona, S., Shaman, J. L., Turitz, A., Wright, J. D., Gyamfi-Bannerman, C., & Metzema, A. (2020). Associations between built environment, neighborhood socioeconomic status, and SARS-CoV-2 infection among pregnant women in New York city. Jama, 324, 390–392.

Evans, N. S., & Masten, F. A. (2012). The phylogenetic Kantorovich-Rubinstein metric for environmental sequence samples. Journal of the Royal Statistical Society, Series C, 61, Article 102672.

Fang, W. and Wahba, S. (2020). Urban density is not an enemy in the coronavirus fight: evidence from China. World Bank Blogs. Available at: https://blogs.worldbank.org/sustainablecities/urban-density-not-enemy-coronavirus-fight/eng-videencia-china.

Fang, H., Wang, L., & Yang, Y. (2020). Human mobility restrictions and the spread of the Novel Coronavirus (2019-nCoV) in China. Journal of Public Economics, 191, Article 104272.

Goldfarb, A., & Tucker, C. E. (2020). Which retail outlets generate the most physical interactions? NBER Working Paper. No. w27042.

Hamidi, S., Ewing, R., & Sabouri, S. (2020). Longitudinal analyses of the relationship between development density and the COVID-19 morbidity and mortality rates: Early evidence from 1,165 metropolitan counties in the United States. Health Place, 64, Article 102378.

Han, Y., Yang, L., Jia, K., Li, J., Feng, S., Chen, W., Zhao, W., & Pereira, P. (2021). Spatial distribution characteristics of the COVID-19 pandemic in Beijing and its relationship with environmental factors. The Science of the Total Environment, 761, Article 144327.

Hishan, S. S., Ramakrishnan, S., Qureshi, M. I., Khan, N., & Al-Kumaim, S. (2020). Pandemic thoughts, civil infrastructure and sustainable development: Five insights from covid-19 across travel lenses. Talent Development and Excellence, 12(3), 169–169.

Hu, M., Roberts, J. D., Azvedo, G. G., & Milner, D. (2021). The role of built and social environmental factors in Covid19 transmission A look at America’s capital city. Sustainable Cities and Society, 65, Article 102580.

Kashem, S. B., Wilson, B., & Van Zandt, S. (2016). Planning for climate adaptation: Evaluating the changing patterns of social vulnerability and adaptation challenges in three coastal cities. Journal of Planning Education and Research, 36(3).

Kimmelman, M. (2020, July 9). New York as a biking city? It could happen. And it should (The New York Times).

Leonard, C. (2012). From the schrodinger problem to the morge-kantorovich problem. Journal of Functional Analysis, 262, 1879–1920.

Li, B., Peng, Y., He, H., Wang, M., & Feng, T. (2011). Built environment and early infection of COVID-19 in urban districts: A case study of Huangzhou. Sustainable Cities and Society, 66, Article 102685.

Liu, R., & Yang, H. (2022). Spatial tessellation of infectious disease spread for epidemic decision support. IEEE Robotics and Automation Letters, 7, 626–633.

Li, X., Zhou, L., Jia, T., Peng, R., Fu, X., & Zou, Y. (2020). Associating COVID-19 severity with environmental factors. The Science of the Total Environment, 761, Article 104272.

Megenah, N. A., & Ghoneim, E. M. (2020). Antivirus-built environment: Lessons learned from Covid-19 pandemic. Sustainable Cities and Society, 61, Article 102356.

Munk, A., & Czado, C. (1998). Nonparametric validation of similar distributions and assessment of goodness of fit. Journal of the Royal Statistical Society - Series B: Statistical Methodology, 60, 223–241.

Nguyen, Q. C., Huang, Y., Kumar, A., Duan, H., Keraz, J. M., Dwivedi, P., Meng, H.-W., Brundish, R. D., & Jay, J. O., Bernardi, M., & Tadzid, T. (2020). Using 164 million Google street view images to derive built environment predictors of COVID-19 cases. International Journal of Environmental Research and Public Health, 17(7), Article 102350.

Ni, K., Bresson, X., Chan, T., & Esedoḡlu, S. (2009). Local histogram based segmentation of Wasserstein distance. International Journal of Computer Vision, 84, 97–117.

Panaretos, V. M., & Zemel, Y. (2019). Statistical aspects of Wasserstein distances. The Annual Annual International Conference on machine learning.

Peng, E., & Shang, X. (2019). The Schrödinger problem and the kantorovich problem. Operations Research, 66, 1603–1624.
Pare, P. E., Liu, J., Beck, C. L., Kirwan, B. E., & Basar, T. (2020). Analysis, estimation, and validation of discrete-time epidemic processes. IEEE Transactions on Control Systems Technology, 28, 79–93.

Rocklov, J., & Sjödin, H. (2020). High population densities catalyse the spread of COVID-19. Journal of Travel Medicine, 27(3).

Rubner, Y., Tomasi, C., & Guibas, L. J. (2000). The Earth Mover’s distance as a metric for image retrieval. International Journal of Computer Vision, 40, 99–121.

Siljander, M., Uusitalo, R., Pellikka, P., Isosomppi, S., & Vapalahti, O. (2022). Spatiotemporal clustering patterns and sociodemographic determinants of COVID-19 (SARS-CoV-2) infections in Helsinki, Finland. Spatial and spatio-temporal epidemiology, 41.

Snow, J. (1855). On the mode of communication of cholera. Cumberland Lodge. U.K.: John Churchill.

Solomon, J., de Goes, F., Peyré, G., Cuturi, M., Butscher, A., Nguyen, A., Du, T., & Guibas, L. (2015). Convolutional Wasserstein distances. ACM Transactions on Graphics, 34, 1–11.

Sun, C., & Zhai, Z. (2020). The efficacy of social distance and ventilation effectiveness in preventing COVID-19 transmission. Sustainable Cities and Society, 62, 102390.