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.

Health & Place 72 (2021) 102694

Contents lists available at ScienceDirect

Health and Place

journal homepage: www.elsevier.com/locate/healthplace

The superspreading places of COVID-19 and the associated built-environment and socio-demographic features: A study using a spatial network framework and individual-level activity data

Jianwei Huang [a], Mei-Po Kwan [b,c,*], Zihan Kan [a]

[a] Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Shatin, Hong Kong, China
[b] Department of Geography and Resource Management and Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Shatin, Hong Kong, China
[c] Department of Human Geography and Spatial Planning, Utrecht University, 3584 CB, Utrecht, the Netherlands

ARTICLE INFO

Keywords:
COVID-19 pandemic
High-risk places
Superspreading places
Individual-level activity data
Spatial network
Built environment
Socio-demographic

ABSTRACT

Previous studies observed that most COVID-19 infections were transmitted by a few individuals at a few high-risk places (e.g., bars or social gathering venues). These individuals, often called superspreaders, transmit the virus to an unexpectedly large number of people. Further, a small number of superspreading places (SSPs) where this occurred account for a large number of COVID-19 transmissions. In this study, we propose a spatial network framework for identifying the SSPs that disproportionately spread COVID-19. Using individual-level activity data of the confirmed cases in Hong Kong, we first identify the high-risk places in the first four COVID-19 waves using the space-time kernel density method (STKDE). Then, we identify the SSPs among these high-risk places by constructing spatial networks that integrate the flow intensity of the confirmed cases. We also examine what built-environment and socio-demographic features would make a high-risk place to more likely become an SSP in different waves of COVID-19 by using regression models. The results indicate that some places had very high transmission risk and suffered from repeated COVID-19 outbreaks over the four waves, and some of these high-risk places were SSPs where most (about 80%) of the COVID-19 transmission occurred due to their intense spatial interactions with other places. Further, we find that high-risk places with dense urban renewal buildings and high median monthly household rent-to-income ratio have higher odds of being SSPs. The results also imply that the associations between built-environment and socio-demographic features with the high-risk places and SSPs are dynamic over time. The implications for better policymaking during the COVID-19 pandemic are discussed.

1. Introduction

The COVID-19 pandemic has become one of the most critical global public health crises since it was designated as a pandemic in March 2020 (World Health Organization, 2020). In the early stage of the pandemic when limited pharmaceutical and vaccine options for treatment and prevention of the disease were available, certain non-pharmaceutical interventions (e.g., social distancing and stay-at-home orders) were considered effective in mitigating the spread of COVID-19 (Gao et al., 2020; Kraemer et al., 2020; Flaxman et al., 2020; Yabe et al., 2020; Hu et al., 2021). A common assumption underlying non-pharmaceutical interventions is that people's mobility can be reduced by these control measures and thus the spread of the virus via face-to-face contact can be mitigated.

As previous studies have found, the transmission of COVID-19 occurs mainly through people's face-to-face interactions in their daily life (especially via respiratory droplets released by infected persons when they cough, sneeze, speak, sing or breathe heavily) (Huang and Kwan, 2021). Further, most people travel to different places in their daily life for various activities (Kwan, 2012), and people's daily mobility outside their homes may not be reduced since they still have to obtain groceries, medicines, essential services, or go to work during a pandemic (Huang et al., 2020). Thus, people's daily mobility is also shaped by various built-environment and socio-demographic features, which may in turn render some places more risky than others (Hutch et al., 2011; Huang et al., 2019; Lai et al., 2020). For instance, using contact tracing data

* Corresponding author. Department of Geography and Resource Management and Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Shatin, Hong Kong, China.
E-mail addresses: Jianwei.Huang@link.cuhk.edu.hk (J. Huang), mpk654@gmail.com (M.-P. Kwan), zihankan@cuhk.edu.hk (Z. Kan).

https://doi.org/10.1016/j.healthplace.2021.102694
Received 20 July 2021; Received in revised form 2 October 2021; Accepted 5 October 2021
Available online 9 October 2021

collected in Hong Kong, Huang et al. (2020), Kan et al. (2021a, b), Kwok et al. (2021), and Yip et al. (2021) found that certain socio-demographic features (e.g., population density, household income, workplace location, and occupation) and built-environment features (e.g., green space, sky view, building density and height, transport nodal accessibility, land use configuration, and street length) significantly affected the spatial patterns of COVID-19 transmission in Hong Kong. Similarly, the same kind of association between built-environment and socio-demographic features with COVID-19 transmission risk has been observed in European countries (Mogi et al., 2021), the U.S. (Raifman and Raifman, 2020; Kim and Kwan, 2021; Huang et al., 2021a), and England (Lee et al., 2021a). Therefore, there is strong evidence from different countries that built-environment and socio-demographic features are contributing factors of COVID-19 transmission. Moreover, studies have observed the existence of spatial nonstationarity in the associations between environmental features and COVID-19 transmission (Kwan, 2021): the associations between environmental features and COVID-19 transmission could change over space.

Although these studies provide a useful foundation for understanding what environmental features would affect the transmission of COVID-19, their conclusions might be misleading due to temporal nonstationarity (Kwan, 2012, 2021). The temporal nonstationarity in health-environment associations means that those associations may change over time. For instance, using data collected from the United States, Hong Kong, and England, studies have found that controlling the pandemic via restricting people's mobility may be effective only for a short period and for certain social groups (Kim and Kwan 2021; Lee et al., 2021a, b). Specifically, people's mobility hardly declined after the early stage of the COVID-19 pandemic (e.g., after June 2020 in the U.S.) despite social distancing measures were still in effect, and poor people experienced a slower decline in mobility in the early stage of the pandemic. Thus, the associations between environmental features and COVID-19 transmission may be different among various waves of COVID-19. In other words, there may be temporal nonstationarity in such associations. Our knowledge about the temporal nonstationarity in the associations between environmental features and COVID-19 transmission is still highly limited to date. It is thus important to explore the temporal nonstationarity in these associations across different waves of COVID-19. The contribution of such analysis is two-fold. On one hand, the new knowledge generated can inform the development and enhance the effectiveness of non-pharmaceutical interventions by targeting specific high-risk places or venues in specific periods. On the other hand, certain features of the environment can be identified and dynamically managed to reduce risky behaviors and thus reduce the transmission of COVID-19.

Further, studies have found that *superspreading events* that occurred at some high-risk places played an important role in driving the transmission of COVID-19: most infections are transmitted by a few individuals at a few high-risk places (e.g., bars or other social gathering venues). These individuals are often called *superspreaders*, who transmit the virus to an unexpectedly large number of people. For instance, Adam et al. (2020) reported that 80% of COVID-19 cases were transmitted by 19% of the infected persons at some high-risk places during the first wave of COVID-19 in Hong Kong. Moreover, Chang et al. (2021) concluded that a small number of *superspreading places* (SSPs) account for a large number of COVID-19 transmissions in the U.S. Thus, in this study, we conceptualize a high-risk place as a place or venue with a high likelihood of COVID-19 transmission through the frequent visits of infected persons. Further, a high-risk place may evolve into a *superspreading place* (SSP) due to its intense spatial interactions with other places, which means that people who are infected at an SSP frequently visit certain places and thus significantly increase the COVID-19 risk of these places. Hence, in this framework, an SSP is a high-risk place with a high ability to spread the disease to many other places. Note that SSPs not only are risky places but also play a pivotal role in spreading the disease due to their superspreading characteristic.

Previous studies have used human mobility data (e.g., smart card data) and spatial networks to examine superspreading places (SSPs). In a spatial network framework, the spatial units (e.g., public transport stations, census tracts, counties) are represented as *nodes*, and the intensity of human movements between the nodes are presented as weighted *edges* that indicate the volume of movements among the nodes (Rizzo et al., 2014; Huang et al., 2019; Kan et al., 2021c; Liu et al., 2021). In a spatial network, the *degree* of a node indicates that the number of edges connected to it; the *strength* of a node refers to the weights of all edges (i.e., the intensity of human movements) connected to it. In past studies, the *rich-club coefficient* was used to assess the extent to which a few powerful nodes (i.e., nodes with high degree and strength) dominate the structure of a network. These powerful nodes will not only form a cohesive cluster among themselves but also maintain their connections with the peripheral nodes (Opsahl et al., 2008). Thus, in the case of COVID-19, the SSPs can be identified using the spatial network framework: the few powerful nodes can be regarded as SSPs. For instance, using daily-aggregated ridership data of public transport in SingaporeChin and Bouffanais (2020) identified the SSPs based on the degree and strength of the spatial network. The study also indicated that a high-risk place may further evolve into an SSP because of intense spatial interactions with other communities (e.g., busy bus interchanges).

While previous studies represent significant advances in the study of SSPs in the COVID-19 pandemic, they did not examine whether the spatial distribution of SSPs varies over time within a city. Further, they did not examine the temporal nonstationarity in the associations between built-environment and socio-demographic features with the SSPs. Thus, this paper seeks to identify the spatial distribution of SSPs in the four waves of the COVID-19 pandemic in Hong Kong and investigate their associations with various built-environment and socio-demographic features. We ask the following research questions: (1) How to identify the SSPs in the four waves of COVID-19 in Hong Kong? (2) Whether the spatial distributions of SSPs vary in the four waves of COVID-19 in Hong Kong? (3) What are the associations between various built-environment and socio-demographic features with the SSPs in the four waves of COVID-19 in Hong Kong (or what built-environment and socio-demographic features would make a place more likely to become an SSP in different COVID-19 waves)?

This study thus seeks to bridge the abovementioned gap by identifying the superspreading places (SSPs) and their associations with various built-environment and socio-demographic features in the four COVID-19 waves in Hong Kong. We first propose an analytical framework that integrates GIS, individual-level activity data, and spatial networks to identify high-risk places and SSPs. The framework is used to identify the high-risk places and SSPs over the four COVID-19 waves in Hong Kong from January 2020 to May 2021. Further, the temporal nonstationarity in the associations between built-environment and socio-demographic features with the high-risk places and SSPs are examined using regression models. Finally, we discuss how the SSPs reflect the social inequality in Hong Kong and their implications for better policymaking during the COVID-19 pandemic.

2. Study area, data and methods

2.1. Study area and dataset

The study area for this research is Hong Kong, which is one of the most densely populated cities in the world. It has an area of 1104 km^2 and an estimated population of 7.5 million in 2020. Thus, the per capita living space in Hong Kong is small (i.e., 14.9 m^2 per person) even when compared to other high-density cities (e.g., Tokyo: 19.5 m^2 per person, and Singapore's 25.0 m^2 per person) (South China Morning Post, 2018). The city is well developed with its built environment characterized by compactivity and connectivity (Kwok et al., 2021). Specifically, Hong Kong follows the pattern of high-rise and high-density development due to limited land resources, where more than 90% of the people are

serviced by the public transport system (Transport Department of Hong Kong, 2020; Yip et al., 2021). These characteristics are helpful for a study that seeks to explore the influences of various environmental features on the transmission of COVID-19.

The COVID-19 dataset used in the study is obtained from the Hong Kong Government's open-data website (https://data.gov.hk). It includes some brief demographic information of all the confirmed cases (e.g., age and gender) from 27 January to 30 May 2021 and the buildings they visited. Note that the confirmed cases are classified into local cases (i.e., cases with clearly local sources) and imported cases (cases infected in a foreign country or city). Fig. 1 shows the four waves of the COVID-19 cases in Hong Kong from Jan 2020 to May 2021: (1) the first wave includes few confirmed cases up to early March of 2020. (2) the second wave happened after a substantial increase in imported cases from March to early May of 2020. (3) the third wave occurred from early July to mid-October of 2020. (4) the fourth wave started in late November 2020 and had over 600 cases linked to the dance club cluster and ended in late May 2021. The study seeks to use the proposed analytical framework to investigate the SSPs in these four COVID-19 waves in Hong Kong. Further, we consider the first and second waves as a continuous wave (i.e., from late January to early May of 2020) and the third wave (i.e., from early July to mid-October 2020) and the fourth wave (i.e., from late Nov 2020 to late May 2021) as two separate waves (see Fig. 1).

The main COVID-19 control measures implemented by the Hong Kong Government from March 2020 up to the end of May 2021 are border control (e.g., a 14-day self-quarantine for in-bound travelers), social distancing (e.g., closure of schools, bars, and clubs, suspension of public services, and "dine-in" restrictions), contact tracing (e.g., in-depth interviews), and location disclosure (e.g., buildings and venues visited by infected people in the 14 days before their infection was confirmed) (Huang et al., 2021b). The government implemented these measures at the outset of the pandemic and they were quite effective during the first and second waves. Then, the measures were relaxed in early June of 2020, which led to the third wave of the outbreak. Thus, the government tightened the measures from the start of the third wave up to the end of the fourth wave. Note that Hong Kong was not locked down during the four COVID-19 waves and people were free to conduct their daily activities and move around, although their travel behaviors and frequencies were changed due to the COVID-19 control measures.

We also used seven types of built-environment data in the study to further investigate the role of built-environment features in the SSPs in the four COVID-19 waves in Hong Kong. We first obtained public transport network data, including data on the Mass Transit Railway (MTR), bus, and ferry services, from the Hong Kong Transport Department. Then, the nodal accessibility of the transport nodes in the public transport system was estimated based on the connectivity matrix of the transport network. The nodal accessibility of each node represents how well the node (i.e., an MTR, bus, or ferry station) is connected with other nodes in the public transport network. Note that we only consider public transport in the study since over 90% of the trips in Hong Kong are made by public transport (Transport Department of Hong Kong, 2020). A building density layer was derived from a building polygon dataset at the 1:1000 scale that the Lands Department provided, and these data were rasterized and resampled to a grid structure with a 100-m spatial resolution. Building density represents the percentage of the area in each 100m × 100m grid cell occupied by buildings and measures the extent of crowdedness due to the number of buildings within a given area (Fig. 2a). A Normalized Difference Vegetation Index (NDVI) layer was derived from SPOT-7 Satellite images (2017) with a spatial resolution of 6m × 6m (Fig. 2b). A land-use dataset was compiled by the Hong Kong Planning Department in 2018, which includes 27 land-use types and with a spatial resolution of 10m × 10m (Fig. 2c). A restaurant dataset was derived from the Openrice website (https://www.openrice.com/), which is the most popular dining guide to help people find restaurants and bars in Hong Kong. The dataset includes various types of social gathering venues (i.e., restaurants and bars) with location information (i.e., latitude and longitude) (Fig. 3a). A markets dataset was provided by the Hong Kong Food and Environmental Hygiene Department. This dataset includes public markets or cooked food markets with location information (i.e., latitude and longitude) (Fig. 3b). An urban renewal dataset was provided by the Hong Kong Urban Renewal Authority (URA). This dataset includes completed URA redevelopment projects with location information (Fig. 3b). Specifically, the dataset shows the buildings in Hong Kong that had been renewed.

In addition, the 2016 census data were used to derive several socio-demographic features in the study. We extracted the data at the Large Street Block Group (LSBG) level via the Census and Statistics Department website (https://geodata.gov.hk/gs/). The LSBG is the smallest planning unit in Hong Kong. The 5034 street blocks (SBs) or village clusters (VCs) in Hong Kong were aggregated by the Census and Statistics Department into 1622 LSBGs (Fig. 4c) to protect personal privacy and data confidentiality. These data include the total population, working population, and the educational, socioeconomic, and housing characteristics of each LSBG. Specifically, they provide information about the age, usual spoken language, educational attainment, occupation, and place of work of the population in each LSBG, as well as the median monthly household income and median monthly household rent-to-income ratio in each LSBG.

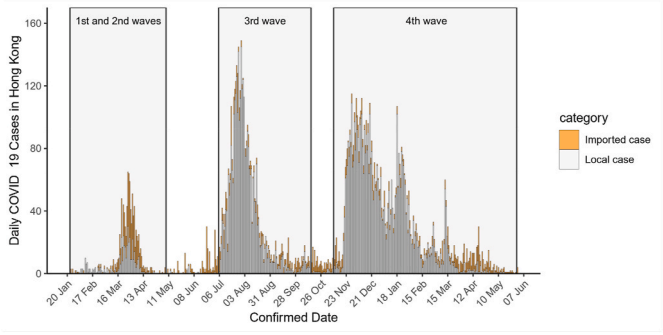

Fig. 1. The daily COVID-19 cases in Hong Kong from January 2020 to May 2021.

2.2. Examining the superspreading places

This study proposes an analytical framework to examine the superspreading places (SSPs) of COVID-19 in Hong Kong. Fig. 4 illustrates the framework that integrates GIS, individual activity data, and spatial networks. The framework first assesses the spatiotemporal distribution of COVID-19 risk and identifies high-risk clusters in Hong Kong based on the space-time kernel density estimation (STKDE) method. Then, the high-risk places are delineated by projecting the high-risk clusters onto continuous 2-D spatial surfaces. We further create the spatial network by representing the high-risk places as nodes and the intensity of spatial interaction (i.e., the movement flows of COVID-19 carriers entailed in their activity patterns) between nodes as weighted edges (or links). Lastly, spatial network properties (i.e., degree and strength) and the rich-club coefficient are used to identify the superspreading places (SSPs).

2.2.1. Detecting and delineating high-risk places based on space-time kernel density estimation (STKDE)

Using the data on the confirmed cases, we first use the space-time kernel density estimation (STKDE) method to generate the spatiotemporal distribution of COVID-19 risk in Hong Kong. Based on the premise that places visited by more infected persons (or confirmed cases) have higher transmission risks, we identify the high-risk clusters and places based on the spatiotemporal distribution of COVID-19 risk. The STKDE is an extension of the traditional kernel density estimation (KDE)

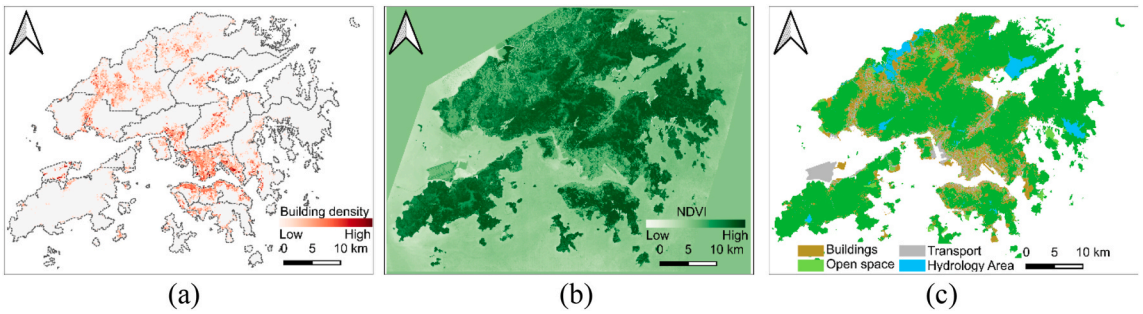

Fig. 2. Spatial distribution of (a) building density, (b) Normalized Difference Vegetation Index (NDVI), and (c) land use in Hong Kong.

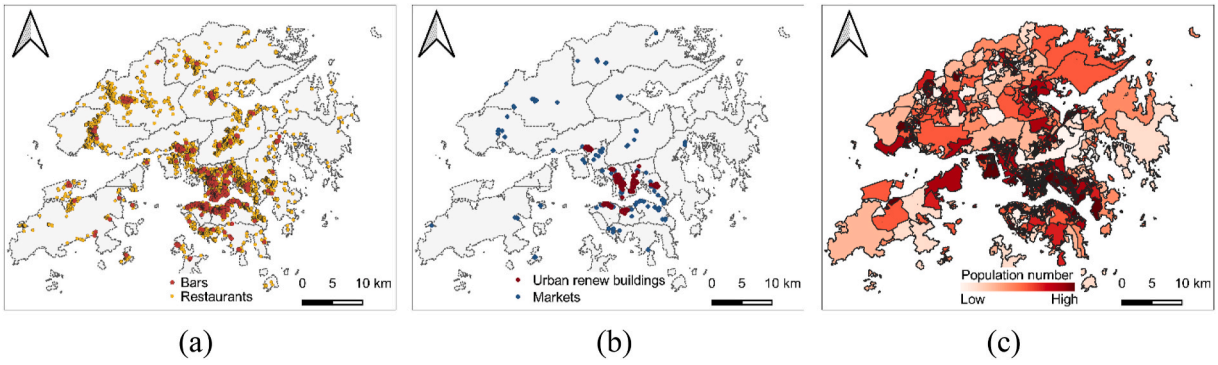

Fig. 3. The spatial distribution of (a) bars and restaurants, (b) renewed buildings and markets, and (c) population in the Large Street Block Groups (LSBGs) in Hong Kong.

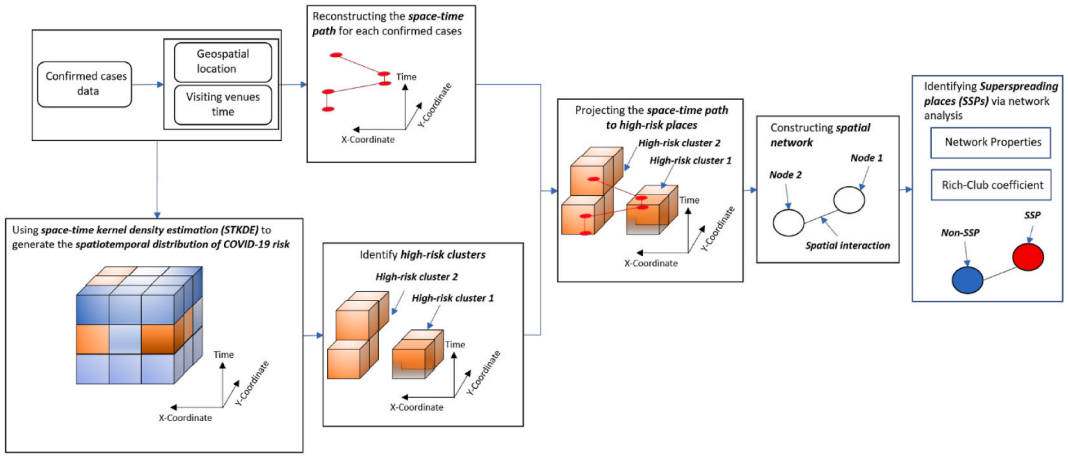

Fig. 4. The framework for identifying COVID-19 superspreading places (SSPs).

method developed by Brunsdon et al. (2007) for detecting the spatio-temporal density of point events (e.g., crime, infectious disease, and traffic accidents) (Nakaya and Yano, 2010; Delmelle et al., 2014; Kang et al., 2018). The space-time density of COVID-19 in the study is estimated by **Equation (1)**:

$$\widehat{f}(x,y,t)=\frac{1}{nh_s^2h_t}\sum_i K_s\left(\frac{x-x_i}{h_s},\frac{y-y_i}{h_s}\right)K_t\left(\frac{t-t_i}{h_t}\right) \quad (1)$$

where (x_i, y_i, t_i) is the space-time location visited by the COVID-19 confirmed cases, $\widehat{f}(x,y,t)$ is the density estimate at location (x,y) at time t. Note that higher density $\widehat{f}$ indicates higher COVID-19 transmission risk. K_s and K_t are the spatial and temporal Epanechnikov kernel function, which is a common method used in ArcGIS (Nakaya and Yano, 2010). Note that the selection of spatial and temporal bandwidth parameters (i.e., h_s and h_t) are critical to STKDE. Thus, we use a space-time Ripley's K function, which can detect the magnitude of clustering at different spatial and temporal bandwidths, to calibrate the optimal spatial and temporal bandwidth parameters (Delmelle et al., 2011; Hohl et al., 2016). In the study, we found that clustering patterns become stronger with increasing spatial bandwidth and decreasing temporal bandwidth. Further, the strongest clustering was found at $h_s = 1000$m, and $h_t = 10$ days. Note that small spatial and temporal bandwidth parameters will generally result in very spiky clusters, while larger parameters may result in oversmoothing (Delmelle et al., 2014). To strike a balance between these two extremes (i.e., the detected clusters being too spiky or oversmoothed), we chose the bandwidth parameters $h_s = 600$m and $h_t = 7$ days, which corresponds to medium magnitude clustering.

Although it would be ideal to create more spatiotemporally detailed surfaces, the spatial and temporal resolutions in the study are 100m × 100m and 1 day because of limitations in the data and computing resources. Further, the high-risk clusters are identified from the density estimates using the top 5% threshold. Finally, to identify high-risk places, we project the detected space-time clusters onto the continuous 2-D (i.e., longitude and latitude) spatial surface of the study area. Thus, the spatial units used in the study is the 100m × 100m cells (used in the STKDE) on this continuous 2-D surface.

2.2.2. Identifying superspreading places (SSP) through spatial network analysis

In this subsection, we used spatial networks to further identify the superspreading places (SSPs). We first infer the movements of the confirmed cases among high-risk places using their activity location records and the derived high-risk places across the four waves. Then, we construct spatial networks across the four waves by representing the high-risk places as nodes, and the intensity of the movements of the confirmed cases between nodes are regarded as weighted edges. Further, we estimate the *degree* and *strength* of these high-risk places in the spatial networks and identify the superspreading places (SSPs) via the *rich-club coefficient* metric. As mentioned in Section 1 above, the degree of a node (i.e., a high-risk place) in the spatial network indicates the number of edges connected to it; the strength of a node refers to the weights of all edges (i.e., the intensity of confirmed cases movements) connected to it. The rich-club coefficient identifies a few powerful nodes (i.e., nodes with high degree and strength), which dominate the structure of the spatial network. In the study, a few powerful nodes will be identified as SSPs.

(1) Constructing spatial networks based on the places visited by the confirmed cases

As mentioned above, we need to use the movement of the confirmed cases to construct the spatial networks, which represent the high-risk places as nodes and the intensity of confirmed cases movements between nodes as weighted edges. To infer the movements of each confirmed case among the high-risk places, we first perform a spatial join analysis between the visit records of each confirmed case and the high-risk places. Note that the high-risk places are generated by using the visit records of the confirmed cases. Hence, 83% of the visit records of the confirmed cases are located in the high-risk places after the spatial join analysis. Thus, each confirmed case's visit records can be denoted as a tuple list of $T = \{(s_1, d_1), (s_2, d_1), \ldots, (s_n, d_m)\}$, where s_i is a high-risk place that a confirmed case visited on day d_i. Then, we derive the movement chains of each confirmed case among the high-risk places as $S = (s_1, s_2, s_3 \ldots s_n)$, where s_i means that the confirmed case has visited high-risk place i, and the tuple (s_i, s_{i+1}) represent the movement between high-risk place i and high-risk place $i+1$. Last, we extract the high-risk place movement chains for all confirmed cases, and the matrix of movements among high-risk places is represented as $M = (s_i, s_j, w)$, where s_i and s_j are the high-risk places i and j, and w is the intensity of flows between s_i and s_j.

After obtaining the movement matrix M, we further construct a spatially weighted network. Note that the network is non-directed since we do not have the time sequence between each movement within one day. In the spatial network, each high-risk place is represented as a node (i.e., s_i is N_i), and the center coordinates (x_i, y_i) of the high-risk place is regarded as the spatial location of the node. Then, we assign a non-directed edge e_{ij} to a pair of nodes (N_i, N_j) depending on whether there was confirmed cases movement between them or not. The weighted W_{ij} of each edge e_{ij} is given by the intensity of flow w between s_i and s_j. Lastly, a weighted non-directed spatial network $G = (N, E, W)$ is constructed.

(2) Exploring the properties of the spatial network

Using the constructed spatial network G, we further estimate the *degree* and *strength* of the high-risk places. Then, the *rich-club coefficient* is used to identify the powerful nodes (i.e., nodes with high degree and strength) in the network, and the powerful nodes would be further identified as superspreading places (SSPs) in the study.

The degree d of a node (i.e., a high-risk place) refers to the number of edges connected to it. Thus, a high-risk place with a higher degree means that it is connected to a larger number of other high-risk places, which means a higher likelihood to spread COVID-19 to other places. The strength s of a node (i.e., a high-risk place) refers to the weights of all edges (i.e., the intensity of flows) connected to it. Thus, the higher strength a high-risk place has, the higher is its capability in spreading COVID-19 to other places.

The *rich-club coefficient* is a metric for detecting whether a few powerful nodes (i.e., nodes with high degree and strength) dominate the structure of a network based on *degree* d and *strength* s. The rich-club coefficient is widely used to explore a two-level network structure (e.g., "rich-poor" network), where a few "rich" nodes will not only form a cohesive cluster among themselves but also maintain their connections with the "poor" nodes (Liu et al., 2021). Nodes with strength and degree greater than a certain value of r are typically considered as "powerful" nodes. Thus, the *rich-club coefficient* of a weighted network can be assessed based on **Equation (2)**:

$$\varphi_w(r) = \frac{W_{>r}}{\sum_{e=1}^{E_{>r}} W_{e,rank}} \tag{2}$$

where $E_{>r}$ refers to the number of edges between the nodes of *degree* greater than r, and $W_{>r}$ is the sum of the *strength* of these edges. $W_{e,rank}$ represents the sum of the strength of the $E_{>r}$, the strongest edges within the whole network. To determine whether a rich-club phenomenon exists in a weighted network, it is necessary to compare the above metric to the same metric on a randomized network using **Equation (3)**:

$$p_w(r) = \frac{\varphi_w(r)}{\varphi_{w,null}(r)} \tag{3}$$

where $\varphi_{w,null}(r)$ is the *rich-club coefficient* assessed on a randomized network based on the null model. When $p_w(r)$ is larger than 1, the network under investigation has a significant rich-club effect. In the study, we use the *rich-club coefficient* to identify the powerful nodes in the weighted non-directed spatial network G, and these powerful nodes would be identified as SSPs.

2.3. The associations between built-environment and socio-demographic features with high-risk places and superspreading places (SSPs)

This subsection explains how we address our last research question: What are the associations between various built-environment and socio-demographic features with the superspreading places (SSPs) in different COVID-19 waves (or what built-environment and socio-demographic features would make a place more likely to become an SSP in different periods)? Previous studies have found that transport nodal accessibility, green space, land-used diversity, population density, and the density of buildings, residential areas, public spaces, restaurants, markets, and workplaces, as well as the occupation and household income of the population in different areas can significantly affect the first to third waves of COVID-19 transmission in Hong Kong (Huang et al., 2020; Kan et al., 2021a, b; Kwok et al., 2021; Yip et al., 2021). However, they did not examine how the associations between those features and SSPs. In addition, it is important to note that schools and medical facilities are not high-risk places in Hong Kong according to the government's daily press briefings (i.e., very few transmissions occurred in these facilities). Therefore, schools and medical facilities were excluded from this study.

The built-environment features selected in the study include nodal

accessibility, green space or vegetation coverage (measured with the Normalized Difference Vegetation Index [NDVI]), land-use diversity, and the density of buildings, residential area, public space, restaurants, markets, and renewed buildings. Table A1 provides a comprehensive description of these selected built-environment features. We derive these built-environment factors within each high-risk place. Specifically, the land-use diversity in each high-risk place is estimated based on the entropy model (Huang et al., 2020; Kan et al., 2021a,b), as **Equation (4)** shows:

$$LUD_i = -\sum_{l=1}^{n} \frac{p_i * \ln p_i}{n} \quad (4)$$

where p_i presents the proportion of ith land-use type, and n is the total number of land-use types.

The socio-demographic features include the density, age, educational attainment, place of work, occupation, median monthly household income, and median monthly household rent-to-income ratio of the population in each LSBG. Table A2 provides a comprehensive description of these selected socio-demographic features. Note that for deriving the age, educational attainment, place of work, and occupation of the population in each LSBG, we first derived the number of persons with certain socio-demographic features in high-risk places by spatial join analysis between LSBGs and high-risk places. Then, all the data within the high-risk places were transformed from the number of persons into ratio variables, which indicate the percentage of the population within the high-risk places with certain sociodemographic characteristics. Further, we derived the median monthly household income and median monthly household rent-to-income ratio in high-risk places by spatial join analysis between LSBGs and high-risk places.

We conduct a regression analysis to examine the association between the high-risk places and SSPs in different COVID-19 waves (i.e., first and second waves, third wave, and fourth wave) with the built-environment and socio-demographic features. First, we focus on the *degree* and *strength* of the high-risk places over different waves (Models 1–6). The dependent variable of Model 1 is the *degree* of high-risk places in the first and second waves, and that of Model 2 is the *strength* of high-risk places in the first and second waves. The dependent variable of Model 3 is the *degree* of high-risk places in the third wave, and that of Model 4 is the *strength* of high-risk places in the third wave. The dependent variable of Model 5 is the *degree* of high-risk places in the fourth wave, and that of Model 6 is the *strength* of high-risk places in the fourth wave. The independent variables of each model are the selected built-environment and socio-demographic features. We choose the ordinary least squares (OLS) regression for Models 1–6, and the Lasso regularization is used to avoid the multicollinearity problem caused by a large number of highly correlated parameters (Kwok et al., 2021).

Second, we investigate the associations between the SSPs and the built-environment and socio-demographic features in different COVID-19 waves (i.e., first and second waves, third wave, and fourth wave) (Models 7–9). The dependent variables in Models 7–9 are the same and indicate whether a high-risk place is an SSP (yes = 1; no = 0). The independent variables of each model are the selected built-environment and socio-demographic features. We use the binary logistic regression models in Models 7–9, and the Lasso regularization is also used to avoid the multicollinearity problem caused by a large number of highly correlated parameters.

3. Results

3.1. Spatial distribution of high-risk places in different waves

This subsection describes the space-time distribution of high-risk places in Hong Kong over the four COVID-19 waves using space-time kernel density estimation (STKDE). Fig. 5 shows the 3D visualization, where the voxels (represented as points) have different opacity and colors based on their COVID-19 risk values (i.e., the density of the places visited by the confirmed cases), which provides a clear display of the patterns of COVID-19 risk in space and time. Each voxel has a spatial resolution of 100m × 100m and a temporal resolution of 1 day. It should be noted that only voxels with the top 5% density values are shown in Fig. 5. The result suggests that the third and fourth waves of the COVID-19 pandemic in Hong Kong were more severe than the first and second waves: the third and fourth waves had a wider spatial extent and longer duration, and the fourth wave lasted the longest (see Fig. 5).

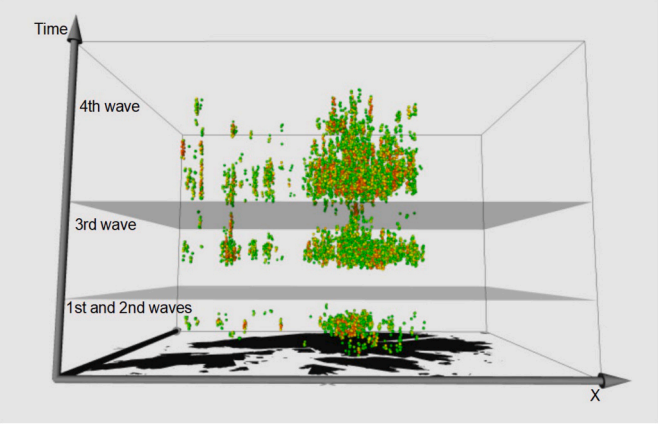

Fig. 5. The spatiotemporal distribution of COVID-19 risk in Hong Kong over the four waves using STKDE (red indicates high risk, while green represents low risk). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

In addition, Fig. 6 shows the spatial distribution of high-risk places over the four waves. Specifically, there were 47, 95 and 94 high-risk places in the first and second waves, the third wave, and the fourth wave. The most risky places in the first and second waves include Centre, Wan Chai, and Causeway Bay (see the blue polygons in Fig. 6). The third wave and the fourth wave had similar spatial patterns, where the most risky places were Sham Shui Po, Yau Tsim Mong, Wong Tai Sin, Kwan Tong, Central, and Wan Chai (see the purple and dark yellow polygons in Fig. 6). The results indicate that the most risky places have suffered more than one COVID-19 outbreak over the four waves. In other words, some places had suffered COVID-19 outbreaks repeatedly over the four waves. One of the reasons is that people's mobility hardly declined after the early stage of the COVID-19 pandemic in Hong Kong. Because people still have to obtain groceries, medicines, essential services, or go to work despite the COVID-19 control measures were still in effect, people's mobility to grocery shopping (e.g., supermarkets, food warehouses, farmers markets, specialty food shops, and pharmacies) in Hong Kong increased 14% and 25% on average during the third wave and fourth wave when compared to the baseline (set as the mean value of mobility from January 3 to February 6, 2020) (Zhang et al., 2021; Google LLC 2021).

3.2. Spatial network construction and characterization

We further identify the superspreading places (SSPs) in Hong Kong based on the identified high-risk places. We first create three spatial networks for the four waves (see Fig. 7): the high-risk places are regarded as nodes, and the intensity of interaction (i.e., the flow of COVID-19 confirmed cases) between nodes are regarded as weighted links. Then, we used the *degree*, *strength* and *rich-club coefficients* to identify the SSPs in the spatial networks of the four COVID-19 waves. Note that the degree of a node in a network refers to the number of connections to other nodes. The strength of a node in a network is the sum of its weights.

Fig. 8 presents the statistical results of the *degree* and *strength* in the

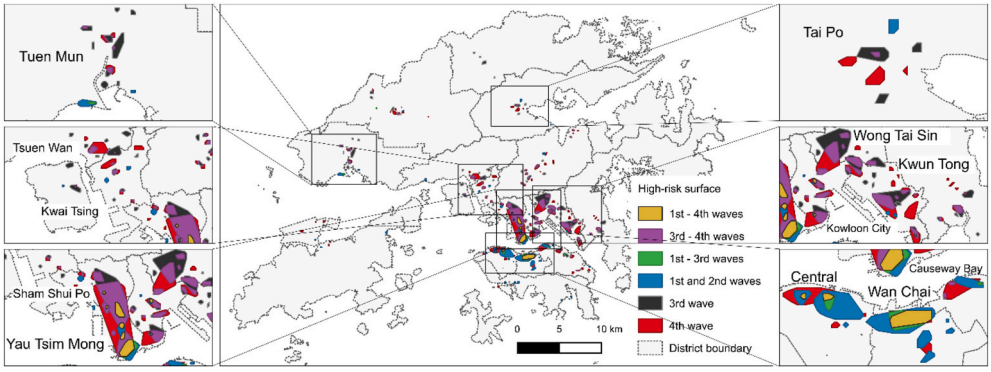

Fig. 6. The spatial distribution of high-risk places in Hong Kong over the four waves.

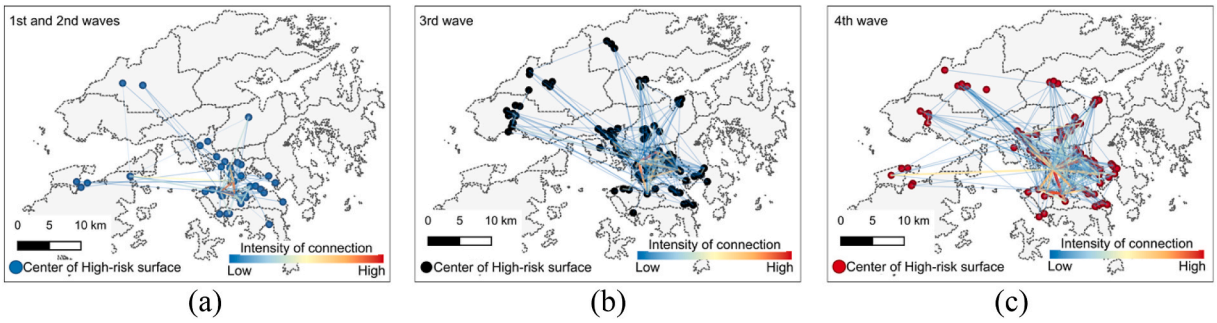

Fig. 7. The spatial networks of the four COVID-19 waves based on the locations visited by the confirmed cases: (a) the first and second waves; (b) the third wave; (c) the fourth wave.

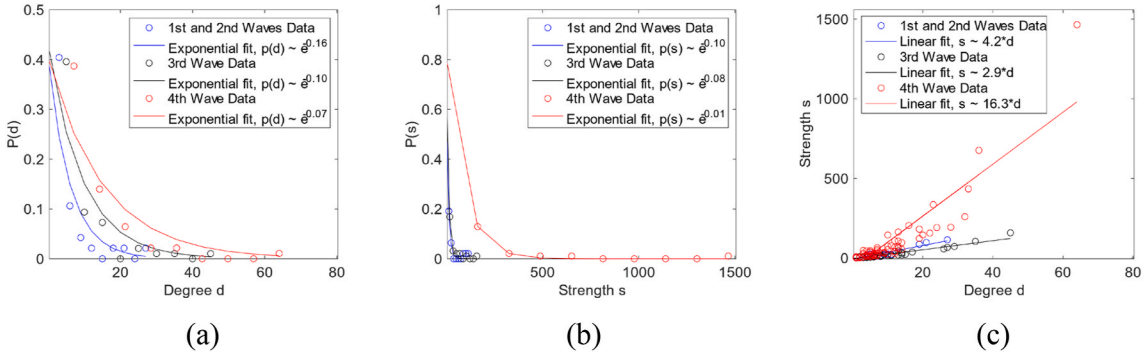

Fig. 8. The degree and strength of the spatial networks: (a)–(b) Probability distributions of degree and strength; (c) Correlations between degree and strength.

three spatial networks. As Fig. 8a and b shows, both the degree and strength of the three spatial networks can be well fitted to the exponential function. In addition, we found that the nodes with a high degree (i.e., $d \geq 6$ for the first and second waves, $d \geq 8$ for the third wave, and $d \geq 10$ for the fourth wave) accounted for only a small proportion of all the nodes: 21%, 22% and 28% for the first and second waves, the third wave and fourth wave, which indicates that relatively few of the high-risk places have high spatial interaction with other places. The nodes with high strength (i.e., $s \geq 11$ for the first and second waves, $s \geq 14$ for the third wave, and $s \geq 60$ for the fourth wave) also accounted for a small proportion of all the nodes: 24%, 22% and 23% for the first and second waves, the third wave and fourth wave, which means that most of the superspreading events occurred in a few high-risk places in Hong Kong over the four waves. These results imply that most of the superspreading events happened at a few of the high-risk places over the four waves, where individuals infected in those superspreading events further infect others in other places through their daily mobility.

We further examine the relationships between the degree and strength of the spatial networks for the four waves (see Fig. 8c). We find that the relationships between the degree and the strength of the spatial networks over the four waves can be fitted to a linear function, which indicates that the spatial interactions between the COVID-19 cases of the high-risk places increase linearly with their degree. The estimated coefficient of the first and second waves is 4.2, which is larger than that of the third wave (i.e., 2.9) but smaller than that of the fourth wave (i.e., 16.3). The result suggests that the COVID-19 cases are more concentrated in a few of the high-risk places in the third wave than in the first and second waves and the fourth wave. In other words, although the number of confirmed cases in the third wave is larger than those of the first and second waves, the increasing cases are more related to the superspreading events in a few of the high-risk places. Further, the fourth wave has the largest estimated linear coefficient (i.e., 16.3), which indicates that the confirmed COVID-19 cases in the fourth wave have higher mobility than the first to third waves, despite the number of

high-risk places in the fourth wave is similar to the third wave. These results imply that a few high-risk places dominated the COVID-19 transmission process in Hong Kong as the number of confirmed cases increased.

3.3. Identifying the spatial distribution of superspreading places (SSPs) in Hong Kong

This subsection applies the *rich-club coefficient* to identify the dominant high-risk places (i.e., the superspreading places [SSPs]) over the four waves. Fig. 9a and b presents the results of the rich-club coefficients, where both $P_w(d)$ and $P_w(s)$ are greater than 1, thus revealing an upward trend over the four waves. The result confirms the existence of an obvious rich-club effect in the spatial transmission of COVID-19 in Hong Kong. Fig. 9a suggests that the degree $d \geq 6$ is an important elbow point (i.e., the degree of nodes change rapidly when $d \geq 6$) for the first and second waves, while $d \geq 8$ and $d \geq 10$ are important for the third and fourth waves. Fig. 9b shows that when $s \geq 11$ is important for the first and second waves, while $s \geq 14$ and $s \geq 60$ are important for the third and fourth waves. Thus, we identify the SSPs over the four waves according to the results. Specifically, nodes with $d \geq 6$ and $s \geq 11$ are identified as SSPs in the first and second waves, nodes with $d \geq 8$ and $s \geq 14$ are identified as SSPs in the third wave, and nodes with $d \geq 10$ and $s \geq 60$ are identified as SSPs in the fourth wave.

In the first and second waves, ten (or 22%) of the 46 high-risk places are identified as SSPs. In the third wave, twenty (or 21%) of the 95 high-risk places are identified as SSPs. In the fourth wave, twenty-five (26%) of the 94 high-risk places are identified as SSPs. Moreover, 75%, 73% and 85% confirmed cases had visited the SSPs in the first and second waves, the third wave, and the fourth wave. The results indicate that most of COVID-19 confirmed cases were infected in a few SSPs, and then the confirmed cases further spread COVID-19 to other places through their daily mobility across the city.

Fig. 10 presents the spatial distribution of the SSPs over the four waves. First, the SSPs in the first and second waves are in Centre, Wan Chai, Causeway Bay, Tsim Sha Tsui and Sham Shui Po. Second, the spatial distribution of the SSPs in the third wave is similar to that in the fourth wave: most of the SSPs are in Sham Shui Po, Yau Tsim Mong, Wong Tai Sin, Kwan Tong, Central, and Wan Chai. Meanwhile, Central has the highest strength (i.e., $s = 114$) and degree (i.e., $d = 27$) in the first and second waves. Sham Shui Po and Yau Tsim Mong have the highest strength (i.e., $s = 230$ for the third wave, while $s = 1465$ for the fourth wave) and degree (i.e., $d = 77$ for the third wave, while $d = 64$ for the fourth wave) in the third and fourth waves.

The results imply that the spatial distribution of SSPs may be similar in different waves of COVID-19 in Hong Kong (e.g., the spatial distribution of SSPs in the third and fourth waves). One possible explanation is that some high-risk places might be the initial SSPs in different COVID-19 waves in Hong Kong (e.g., Central, Wan Chai, Causeway Bay, Tsim Sha Tsui and Sham Shui Po). These initial SSPs played a pivotal role in spreading COVID-19 since the disease would spread from the initial SSPs to other places due to people's daily mobility. Note that people's mobility hardly declined after the early stage of COVID-19 in Hong Kong. Further, the spatial distribution of people's daily mobility in a city tends to be quite regular over a certain period (e.g., hours, days, and weeks) (Sevtsuk and Ratti, 2010; Huang et al., 2019). Thus, people's regular daily mobility would further shape the similar spatial distribution of high-risk places and SSPs in different waves of COVID-19 in Hong Kong.

3.4. The associations between built-environment features with the high-risk places and superspreading places (SSPs)

As mentioned in Section 1, people's daily mobility is shaped by the built-environment and socio-demographic features, which may further shape the spatial distribution of high-risk places and SSPs. Thus, we investigate the associations between various built-environment and socio-demographic features with the high-risk places and superspreading places (SSPs). Table 1 shows the regression results on the associations between the degree and strength of the high-risk places with selected built-environment and socio-demographic features in the study area. The dependent variable for Models 1, 3 and 5 is the degree of the high-risk places, whereas the dependent variable for Models 2, 4 and 6 is the strength of the high-risk places.

The results of Models 1–6 reveal that most of the built-environment features have significant associations with the degree and strength of high-risk places in the four waves. Specifically, the density of renewed buildings has the largest contribution to the degree and strength of the high-risk places in the four waves. Further, nodal accessibility, green space (NDVI) and restaurant density have positive associations with the degree and strength of the high-risk places in the third and fourth waves, while they do not have significant associations with the degree and strength of the high-risk places in the first and second waves. Restaurant density has a positive association with the degree and strength of the high-risk places in the first and second waves and the third wave, while it does not have a significant association with the degree and strength of the high-risk places in the fourth wave. Land-use diversity has a significant positive association with the degree and strength of the high-risk places in the four waves, but it does not have a significant association with the degree of the high-risk places in the first and second waves.

In addition, the results of Models 1–6 also reveal that some of the socio-demographic features have significant associations with the degree and strength of the high-risk places in the four waves. Specifically, the median monthly household rent-to-income ratio has a significant

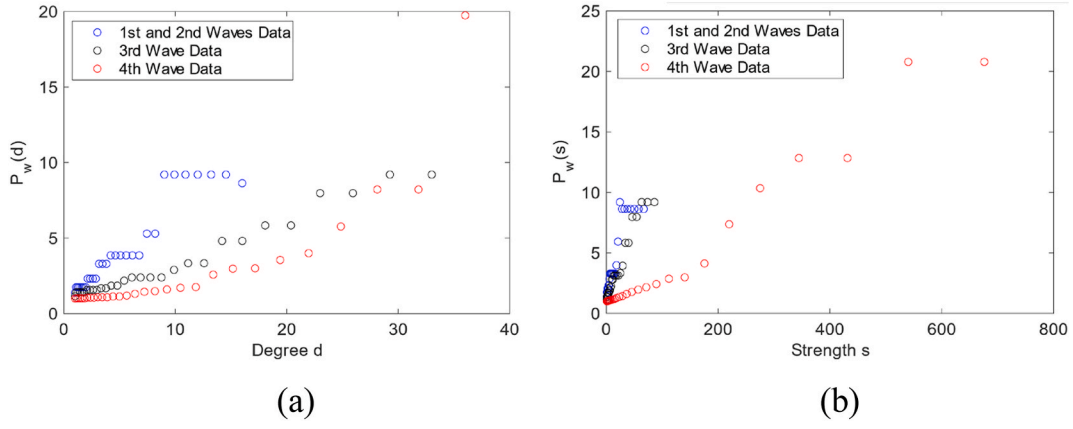

Fig. 9. Rich-club coefficients over the four waves: (a) when $r = d$; (b) when $r = s$.

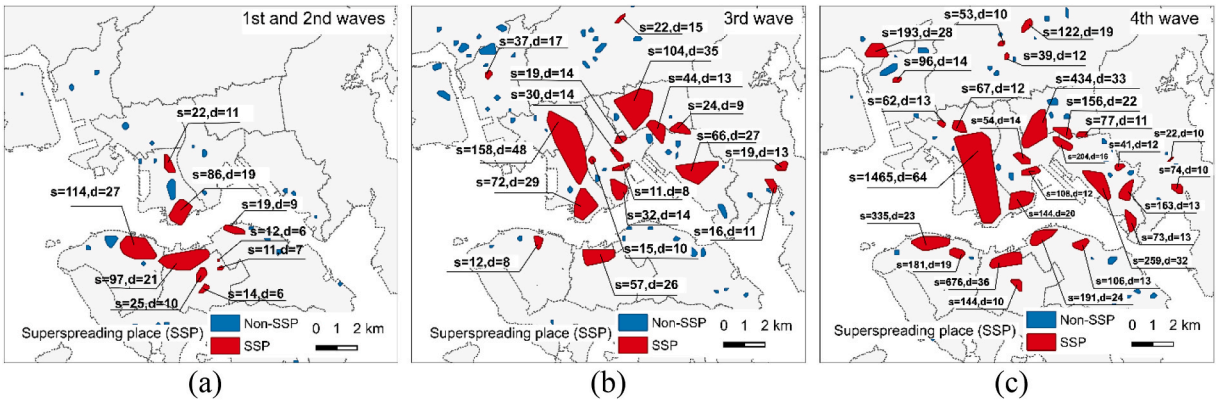

Fig. 10. Spatial distribution of the SSPs in the four waves: (a) the first and second waves; (b) the third wave; (c) the fourth wave.

Table 1
Results of the regression models on the association between built-environment features with high-risk place degree and strength over the four waves in Hong Kong.

	First and second waves		Third wave		Fourth wave	
	Model 1[(a)]	Model 2[(b)]	Model 3[(a)]	Model 4[(b)]	Model 5[(a)]	Model 6[(b)]
Nodal accessibility	-	-	0.152**	0.134***	0.142**	0.167***
Building density	−0.962	-	0.215**	-	0.155*	-
Green space	-	-	0.221**	0.384***		0.128***
Land use diversity	0.029	0.247**	0.410***	0.415***	0.565***	0.851***
Residential density	0.238*	0.356***	0.183***	0.232***	-	-
Public space density	0.223*	-	-	-	-	0.116***
Restaurant density	-	-	0.112**	0.175***	0.048	0.210***
Market density	−0.270*	−0.086	−0.136**	-	-	−0.310***
Urban renewal building density	1.231***	1.875***	0.469***	0.622***	0.231**	0.323***
Population density	-	0.267***	-	0.142***	-	0.069***
Age group 1	-	-	-	-	-	-
Age group 2	-	-	-	-	-	-
Age group 3	-	-	-	-	-	-
Age group 4	-	-	-	-	-	-
Age group 5	-	-	-	0.172***	-	0.192***
Cantonese speaking population	-	-	-	-	-	-
Education group 1	-	-	-	-	-	-
Education group 2	-	-	-	-	-	-
Education group 3	-	-	-	-	-	-
Workplace 1	-	-	-	-	-	-
Workplace 2	-	−0.148**	-	0.502***	-	0.060**
Workplace 3	-	-	-	-	-	0.462***
Workplace 4	-	-	-	-	-	0.154***
Occupation 1	-	-	-	-	-	-
Occupation 2	-	-	-	-	-	-
Occupation 3	-	-	-	-	-	-
Occupation 4	-	-	-	-	-	-
Occupation 5	-	-	-	-	0.338***	0.595***
Occupation 6	-	−0.162	-	-	-	-
Occupation 7	-	-	-	-	-	-
Occupation 8	-	-	-	-	-	-
Occupation 9	−0.209	-	-	-	−0.121**	−0.179***
Median monthly household income	-	-	-	-	-	-
Median monthly household rent-to-income ratio	0.344***	0.309***	0.430***	0.557***	0.227***	0.528***
Intercept	0.718***	1.159***	1.229***	1.537***	1.693***	3.159***
Log-Likelihood	−92.104	−154.737	−219.045	−357.298	−261.707	−1197.82
AIC	208.209	329.474	466.090	734.593	553.413	2439.644
Observations	47	47	95	95	94	94

Notes: *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$; Standard errors in parenthesis; (a) Dependent variable: the degree of high-risk places; (b) Dependent variable: the strength of high-risk places.

positive association with the degree and strength of the high-risk places in the four waves. Population density has a significant positive association with the strength of the high-risk places in the four waves, while it does not have a significant association with the degree of the high-risk places in the four waves. Age Group 5 (i.e., the percentage of the population aged over 64 in high-risk places) has a significant positive association with the strength of the high-risk places in the third and fourth waves. Workplace 2 (i.e., the percentage of the working population in high-risk places with workplaces in Kowloon) has a negative association with the strength of high-risk places in the first and second waves, while this association is significant but positive in the third and fourth waves. Workplace 3 and Workplace 4 (i.e., the percentage of the working population in high-risk places with workplaces in the New Territories or no fixed workplace) have positive associations with the strength of high-risk places in the fourth wave. Occupation 5 (i.e., the percentage of the working population with service and sales workers in high-risk places)

has a significant positive association with the strength and degree of high-risk places in the fourth wave, while Occupation 9 (i.e., the percentage of the working population with skilled agricultural and fishery workers or occupations not classifiable in high-risk places) has a negative association in the fourth wave.

We also examine the associations between the built-environment and socio-demographic features with the odds of being an SSP for the four COVID-19 waves. Table 2 shows the results of the autologistic regression models (i.e., Models 7–9). The results suggest that the high-risk places with dense renewed buildings and higher median monthly domestic household rent-to-income ratio have higher odds of being SSPs over the four waves. The results of Models 8 and 9 also indicate that nodal accessibility, land-use diversity and Workplace 2 (i.e., the percentage of the working population in high-risk places with workplaces in Kowloon) have significant positive associations with the odds of being an SSP in the third and fourth waves, while the density of markets has a negative one.

These results indicate that the density of renewed buildings and the median monthly household rent-to-income ratio played important roles in the transmission of COVID-19 in Hong Kong. Note that these two features were ignored by most previous studies of COVID-19 in Hong Kong (Huang et al., 2020; Kan et al., 2021a, b; Kwok et al., 2021; Yip et al., 2021). Specifically, the density of renewed buildings and the median monthly household rent-to-income ratio significantly increase the likelihood of COVID-19 superspreading events at a high-risk place, from which COVID-19 can then spread to other places. Thus, urban renewal areas with a high median monthly household rent-to-income ratio have a higher likelihood of being SSPs than other high-risk places in Hong Kong. One of the reasons for this is that there are many old apartment buildings (or tenements) in the urban renewal areas, and most of them are rented by people who have a low income. Thus, these old buildings are potential hotbeds of COVID-19 transmission because of their cramped quarters, faulty piping, and poor ventilation (Ho et al., 2012). Further, those areas are also one of the most diverse (e.g., diverse races and nationalities) and densely packed neighborhoods known for their high concentrations of subdivided flats or apartments where the residents are among the poorest people in Hong Kong (The New York Times, 2021).

Table 2
Results of the autologistic regression model on the association between built-environment features and the odds of being superspreading places (SSPs) in the four COVID-19 waves in Hong Kong.

	First and second waves	Third wave	Fourth wave
	Model 7[(a)]	Model 8[(a)]	Model 9[(a)]
Nodal accessibility	-	1.133**	1.278**
Building density	-	-	0.864
Green space	-	-	-
Land use diversity	-	1.788**	2.188**
Residential density	-	-	-
Public space density	-	0.055	-
Restaurant density	-	-	0.717
Market density	-	−0.672*	−0.916*
Urban renewal building density	1.238***	1.525***	1.631***
Population density	-	-	1.407**
Age group 1	-	-	-
Age group 2	-	-	-
Age group 3	-	-	-
Age group 4	-	-	-
Age group 5	-	0.012	-
Cantonese speaking population	-	-	-
Education group 1	-	-	-
Education group 2	-	-	-
Education group 3	-	-	-
Workplace 1	-	-	-
Workplace 2	-	0.993**	1.34**
Workplace 3	-	-	−0.429
Workplace 4	-	-	-
Occupation 1	-	-	-
Occupation 2	-	-	-
Occupation 3	-	-	-
Occupation 4	-	-	0.603
Occupation 5	-	-	-
Occupation 6	-	-	-
Occupation 7	-	-	-
Occupation 8	-	-	-
Occupation 9	-	-	-
Median monthly household income	-	-	-
Median monthly household rent-to-income ratio	2.611***	0.570*	1.643***
Intercept	−2.982***	−1.689***	−2.161***
Log-Likelihood	−11.091	−26.528	−20.816
AIC	32.182	71.056	67.631
Observations	47	95	94

Notes: *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$; Standard errors in parenthesis; (a) Dependent variable is SSP_i (1: the high-risk place is SSP; 0: otherwise).

In addition, the results suggest that nodal accessibility is not a significant factor in the first and second waves of COVID-19 in Hong Kong. The results are different from the findings of previous studies, which found that higher nodal accessibility significantly increases the incidence rate and venue-based transmission risk of COVID-19 (based on the density of venues visited by infected persons) (Huang et al., 2020; Kan et al., 2021a). One possible explanation for such different results is the use of different areal units: the two previous studies used aggregate data at the level of Tertiary Planning Units or Large Street Block Groups, while this study used a grid structure composed of 100m × 100m cells in the space-time kernel density estimation (STKDE). Moreover, the study observes that building density, green space, land-used diversity, residential density, market density, restaurant density, population density, age groups, occupation, and workplace location have associations with the degree or strength of high-risk places in different waves, while people's education level does not have the associations with the degree or strength of high-risk places in different waves, which is in line with the conclusions from previous studies (Huang et al., 2020; Kan et al., 2021a, b; Kwok et al., 2021).

Lastly, the results also indicate that different built-environment and socio-demographic features played different roles in different waves of the COVID-19 pandemic in Hong Kong, which means that temporal nonstationarity exists in the relationships between environmental and socio-demographic factors and COVID-19 transmission. Further, compared to the first and second waves, the third and fourth waves have similar patterns of associations between built-environment and socio-demographic features with SSPs. For example, nodal accessibility, land-use diversity, and Workplace 2 (i.e., the percentage of the working population in high-risk places with workplace in Kowloon) can significantly increase the odds of being an SSP in the third or fourth waves, while they do not have that association with SSPs in the first and second waves. One of the reasons for this is that people's mobility hardly declined after the first and second waves of COVID-19 (e.g., after June 2020) in Hong Kong, despite social distancing measures were still in effect (Lee et al., 2021b). Thus, the associations between environmental features and COVID-19 transmission change over different waves of COVID-19 in Hong Kong. These results imply that the associations between built-environment and socio-demographic features with the high-risk places and SSPs are dynamic over space and time. Specifically, those associations may change under different non-pharmaceutical interventions, and the transmission risk of COVID-19 in different places may increase or decrease over time (Kwan, 2021).

4. Conclusion and discussion

In this study, we propose an analytical framework that integrates GIS, individual activity data, and spatial networks to identify the high-risk places and superspreading places (SSPs) in the first four waves of COVID-19 in Hong Kong. With this framework, we first identify the high-risk clusters and places by using the space-time kernel density

estimation (STKDE) method based on the activity locations of COVID-19 confirmed cases. Then, we reconstruct the spatial movements of the confirmed cases among the high-risk places and create spatial networks by representing the high-risk places as nodes and the movement flows of the COVID-19 confirmed cases between nodes as weighted links. Spatial network properties (i.e., degree and strength) and the rich-club coefficient are used to identify the SSPs. Lastly, the associations between various built-environment and socio-demographic features with the high-risk places and SSPs are examined using regression models. The main findings and implications of the study are summarized as follows.

The results reveal that some places in Centre, Wan Chai, Causeway Bay, Tsim Sha Tsui and Sham Shui Po have very high transmission risk and suffered from COVID-19 outbreaks over the four waves. The results further show that these high-risk places played a significant role in spreading COVID-19 to other communities due to the higher intensity of their spatial interactions (i.e., degree and strength) with other places. Thus, some of the high-risk places further evolve into SSPs where most of the COVID-19 transmissions occurred (i.e., around 80% of the COVID-19 confirmed cases had visited the SSPs in the four waves). Lastly, the regression models indicate that certain built-environment and socio-demographic features have significant associations with high-risk places and SSPs. Specifically, high nodal accessibility, land-use diversity, the density of urban renewal areas, and a high median monthly household rent-to-income ratio are associated with higher odds of being SSPs.

The study is important because it is among the first study that examined SSPs using up-to-date data that cover all four COVID-19 waves in Hong Kong. The study also provides strong evidence that temporal nonstationarity exists in research on the environmental determinants of COVID-19 transmission. This corroborates the findings in other studies that the spatial distributions of COVID-19 are different between the first and second waves in other regions of the world, and thus the associations between built-environment and socio-demographic features with COVID-19 transmission could be different between different COVID-19 waves (Golinelli et al., 2021; Marí-Dell'Olmo et al., 2021; Shim et al., 2021). Our results based on the data of all four waves further reveal that the spatial distributions of COVID-19 in the third and fourth waves are similar in Hong Kong. One of the reasons for this is that people's mobility hardly declined after the early stage of COVID-19 (e.g., after June 2020) despite social distancing measures were still in effect. This led to repeated COVID-19 outbreaks in some of the high-risk places in the third and fourth waves. And some of these high-risk places further became SSPs and played a significant role in spreading the disease to other communities due to the high intensity of their spatial interaction (i.e., degree and strength) with other places.

Further, the findings have several important implications for policymakers when formulating non-pharmaceutical interventions to control the transmission of COVID-19. First, our results reveal that around 80% of COVID-19 confirmed cases visited a few specific places (e.g., Wan Chai, Yau Tsim Mong, Sham Shui Po, Wong Tai Sin and Kwun Tong) in their daily life, and those places were the high-risk places that experienced repeated COVID-19 outbreaks. Specifically, these high-risk places have intense spatial interactions with other communities, which render them to evolve into SSPs through spreading the disease to many other places. Our findings thus suggest that policymakers should put more resources to target these high-risk places to improve the ability to control and prevent COVID-19 outbreaks (e.g., contact tracing, testing and vaccination).

Second, our results indicate that high-risk places with a high density of urban renewal buildings and a high median monthly household rent-to-income ratio are associated with higher odds of being SSPs. Specifically, some social groups may be doubly disadvantaged if they live in these areas. For instance, people with disadvantage socioeconomic status (e.g., unemployed, low income or a high median monthly household rent-to-income ratio) do not have many options regarding which trips to make or forego. Hence, they may have to expose to high-risk environments even if there are COVID-19 outbreaks in their communities. Moreover, they may still expose to high-risk environments even if they can stay at home since their old and dilapidated apartment buildings are potential hotbeds of COVID-19 transmission (Morawska et al., 2020; Nature, 2021). Thus, our findings suggest that policymakers should be aware that certain social groups in high-risk places may be doubly disadvantaged when formulating measures to control the transmission of COVID-19.

Third, our findings also call for revisiting social vulnerability metrics by integrating the built environment and people's daily mobility during the pandemic. The existing social vulnerability metrics (e.g., the Social Vulnerability Index and the Social Health Index) have successfully help policymakers identified vulnerable populations and communities (Cutter, 1996). However, studies on vulnerable populations have concentrated on people's socioeconomic status and demographic factors (Tate, 2013; Oulahen et al., 2015). Our findings indicate that people may be doubly disadvantaged due to their residential environments (e.g., old and subdivided apartments in urban renewal areas with high COVID-19 transmission risk) and socio-demographic features (e.g., unemployed, low income or high median monthly household rent-to-income ratio). Hence, social vulnerability metrics should be integrated with the built environment and people's daily mobility because certain types of vulnerable populations might not be exposed to the same level of COVID-19 risks.

Meanwhile, the study has several limitations. First, there is much potential to further extend our study on the SSPs across the four waves of COVID-19 in Hong Kong. For instance, a comparative analysis of SSPs across the multiple waves of COVID-19 in different cities (e.g., Seoul, London, New York, Tokyo, and so on) could be conducted if similar data are available in these cities. Thus, future studies would reveal more systematic urban social inequity (e.g., race, gender, children, the poor, the aging people) due to the multiple waves of COVID-19 across the world if comparative studies among different cities could be conducted. Finally, the results of STKDE used in the study are determined by the sizes of the spatial and temporal bandwidths. Thus, the results of STKDE may be different if the sizes of the bandwidths are changed. Although this study adopts the well-known optimal bandwidth selection methods (i.e., space-time Ripley's K function) for the STKDE, further investigation is necessary to examine the impact of bandwidth size on STKDE results.

Declaration of competing interest

The author declares no competing financial interests.

Acknowledgments

The authors would like to thank the Hong Kong Department of Health for its kind support in providing public access to the individual-level data used in this research. This research was supported by grants from the Hong Kong Research Grants Council (General Research Fund Grant no. 14605920, 14611621; Collaborative Research Fund Grant no. C4023-20GF) and a grant from the Research Committee on Research Sustainability of Major Research Grants Council Funding Schemes of the Chinese University of Hong Kong. Zihan Kan was supported by an RGC Postdoctoral Fellowship from the Research Grants Council of Hong Kong (PDFS2021-4S08).

Appendix A

Table A1
Descriptions of the selected built-environmental features of the high-risk places

	Data source	Name	Descriptions
Built-environmental features	Public transport network	*Nodal accessibility*	The nodal accessibility of a high-risk place is derived as the average nodal accessibility of all the public transit stations in the high-risk place.
	Building polygons	*Building density*	The building density in each high-risk place is estimated by dividing the building area by the area of the high-risk place.
	SPOT-7 Satellite images	*NDVI*	The area of green spaces in each high-risk place is estimated using the sum of NDVI within the high-risk place.
	Land-use data	*Land use diversity*	The land-use diversity variable represents the land-use mix for each high-risk place, which is calculated based on the entropy model.
		Residential density	The residential density in each high-risk place is estimated by dividing the residential land use area by the area of the high-risk place.
		Public space density	The public space density in each high-risk place is estimated by dividing public land-use area by the area of the high-risk place.
	Openrice website	*Restaurant density*	The density of restaurants in each high-risk place is estimated by dividing the number of restaurants within the place by the area of the high-risk place.
	Markets data	*Market density*	The market density in each high-risk place is estimated by dividing the number of markets within the place by the area of the high-risk place.
	Urban renewal dataset	*Urban renewal building density*	The urban renewal building density in each high-risk place is estimated by dividing the building area under urban renewal by the area of the high-risk place.

Table A2
Descriptions of the selected social-demographic features of the high-risk places

	Sub-category	Name	Descriptions
Socio-demographic features	Demographic features	*Population density*	The number of people per square kilometer.
		Age group 1	The percentage of the population aged below 15.
		Age group 2	The percentage of the population aged between 15 and 24.
		Age group 3	The percentage of the population aged between 25 and 44.
		Age group 4	The percentage of the population aged between 45 and 64.
		Age group 5	The percentage of the population aged over 64.
		Cantonese speaking population	The percentage of people whose usual spoken language is Cantonese.
	Educational features	*Education group 1*	The percentage of people whose highest education level is primary education and below.
		Education group 2	The percentage of people whose highest education level is secondary education.
		Education group 3	The percentage of people whose highest education level is post-secondary education.
	Economic features	*Work place 1*	The percentage of people whose workplaces are on Hong Kong Island.
		Work place 2	The percentage of peoples whose workplaces are in Kowloon.
		Work place 3	The percentage of people whose workplaces are in the New Territories.
		Work place 4	The percentage of people who have no fixed workplace in Hong Kong.
	Occupation features	*Occupation 1*	The percentage of managers and administrators in the working population.
		Occupation 2	The percentage of professionals in the working population.
		Occupation 3	The percentage of associate professionals in the working population.
		Occupation 4	The percentage of clerical support workers in the working population.
		Occupation 5	The percentage of service and sales workers in the working population.
		Occupation 6	The percentage craft and related workers in the working population.
		Occupation 7	The percentage of plant and machine operators and assemblers in the working population.
		Occupation 8	The percentage of people with elementary occupations in the working population.
		Occupation 9	The percentage of the working population with skilled agricultural and fishery jobs or occupations not classifiable.
	Household features	*Median monthly household income*	The median monthly household income.
		Median monthly domestic household rent-to-income ratio	The median monthly domestic household rent-to-income ratio.

References

Adam, D.C., Wu, P., Wong, J.Y., Lau, E.H., Tsang, T.K., Cauchemez, S., Leung, G.M., Cowling, B.J., 2020. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. Nat. Med. 26, 1714–1719. https://doi.org/10.1038/s41591-020-1092-0.

Brunsdon, C., Corcoran, J., Higgs, G., 2007. Visualising space and time in crime patterns: a comparison of methods. Comput. Environ. Urban Syst. 31 (1), 52–75. https://doi.org/10.1016/j.compenvurbsys.2005.07.009.

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

Chin, W.C.B., Bouffanais, R., 2020. Spatial super-spreaders and super-susceptibles in human movement networks. Sci. Rep. 10, 18642. https://doi.org/10.1038/s41598-020-75697-z.

Cutter, S.L., 1996. Vulnerability to environmental hazards. Prog. Hum. Geogr. 20 (4), 529–539. https://doi.org/10.1177/030913259602000407.

Delmelle, E., Delmelle, E.C., Casas, I., Barto, T., 2011. HELP: a GIS-based health exploratory analysis tool for practitioners. Appl. Spat. Anal. Policy 4 (2), 113–137. https://doi.org/10.1007/s12061-010-9048-2.

Delmelle, E., Dony, C., Casas, I., Jia, M., Tang, W., 2014. Visualizing the impact of space-time uncertainties on dengue fever patterns. Int. J. Geogr. Inf. Sci. 28 (5), 1107–1127. https://doi.org/10.1080/13658816.2013.871285.

Flaxman, S., Mishra, S., Gandy, A., Unwin, H.J.T., Mellan, T.A., Coupland, H., Whittaker, C., Zhu, H., Berah, T., Eaton, J.W., Monod, M., 2020. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature 584, 257–261. https://doi.org/10.1038/s41586-020-2405-7.

Gao, S., Rao, J., Kang, Y., Liang, Y., Kruse, J., Dopfer, D., Sethi, A.K., Reyes, J.F.M., Yandell, B.S., Patz, J.A., 2020. Association of mobile phone location data indications of travel and stay-at-home mandates with COVID-19 infection rates in the US. JAMA Netw. Open 3 (9), e2020485. https://doi.org/10.1001/jamanetworkopen.2020.20485.

Golinelli, D., Lenzi, J., Adja, K.Y.C., Reno, C., Sanmarchi, F., Fantini, M.P., Gibertoni, D., 2021. Small-scale spatial analysis shows the specular distribution of excess mortality between the first and second wave of the COVID-19 pandemic in Italy. Publ. Health 194, 182–184. https://doi.org/10.1016/j.puhe.2021.03.008.

Google, L.L.C., 2020. Google COVID-19 community mobility reports. https://www.google.com/covid19/mobility/. (Accessed 29 September 2021). Accessed.

Ho, D.C.W., Yau, Y., Poon, S.W., Liusman, E., 2012. Achieving sustainable urban renewal in Hong Kong: strategy for dilapidation assessment of high rises. J. Urban Plann. Dev. 138 (2), 153–165. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000104.

Hohl, A., Delmelle, E., Tang, W., Casas, I., 2016. Accelerating the discovery of space-time patterns of infectious diseases using parallel computing. Spat. Spatiotemporal Epidemiol. 19, 10–20. https://doi.org/10.1016/j.sste.2016.05.002.

Hu, S., Xiong, C., Yang, M., Younes, H., Luo, W., Zhang, L., 2021. A big-data driven approach to analyzing and modeling human mobility trend under non-pharmaceutical interventions during COVID-19 pandemic. Transport. Res. C Emerg. Technol. 124, 102955. https://doi.org/10.1016/j.trc.2020.102955.

Huang, J., Kwan, M.-P., 2021. Uncertainties in the assessment of COVID-19 risk: a Study of people's exposure to high-risk environments using individual-level activity data. Ann. Assoc. Am. Geogr. https://doi.org/10.1080/24694452.2021.1943301.

Huang, J., Liu, X., Zhao, P., Zhang, J., Kwan, M.-P., 2019. Interactions between bus, metro, and taxi use before and after the Chinese Spring Festival. ISPRS Int. J. Geo-Inf. 8 (10), 445. https://doi.org/10.3390/ijgi8100445.

Huang, J., Kwan, M.-P., Kan, Z., Wong, M.S., Kwok, C.Y.T., Yu, X., 2020. Investigating the relationship between the built environment and relative risk of COVID-19 in Hong Kong. ISPRS Int. J. Geo-Inf. 9 (11), 624. https://doi.org/10.3390/ijgi9110624.

Huang, X., Lu, J., Gao, S., Wang, S., Liu, Z., Wei, H., 2021a. Staying at home is a privilege: evidence from fine-grained mobile phone location data in the United States during the COVID-19 pandemic. Ann. Assoc. Am. Geogr. https://doi.org/10.1080/24694452.2021.1904819.

Huang, J., Kwan, M.-P., Kim, J., 2021b. How culture and sociopolitical tensions might influence people's acceptance of COVID-19 control measures that use individual-level georeferenced data. ISPRS Int. J. Geo-Inf. 10 (7), 490. https://doi.org/10.3390/ijgi10070490.

Hutch, D.J., Bouye, K.E., Skillen, E., Lee, C., Whitehead, L., Rashid, J.R., 2011. Potential strategies to eliminate built environment disparities for disadvantaged and vulnerable communities. Am. J. Publ. Health 101 (4), 587–595. https://doi.org/10.2105/AJPH.2009.173872.

Kan, Z., Kwan, M.-P., Wong, M.S., Huang, J., Liu, D., 2021a. Identifying the space-time patterns of COVID-19 risk and their associations with different built environment features in Hong Kong. Sci. Total Environ. 772, 145379. https://doi.org/10.1016/j.scitotenv.2021.145379.

Kan, Z., Kwan, M.-P., Huang, J., Wong, M.S., Liu, D., 2021b. Comparing the space-time patterns of high-risk areas in different waves of COVID-19 in Hong Kong. Trans. GIS. https://doi.org/10.1111/tgis.12800.

Kan, Z., Kwan, M.-P., Tang, L., 2021c. Ripley's K function for network constrained flow data. Geogr. Anal. https://doi.org/10.1111/gean.12300.

Kang, Y., Cho, N., Son, S., 2018. Spatiotemporal characteristics of elderly population's traffic accidents in Seoul using space-time cube and space-time kernel density estimation. PLoS One 13, e0196845. https://doi.org/10.1371/journal.pone.0196845.

Kim, J., Kwan, M.-P., 2021. The impact of the COVID-19 pandemic on people's mobility: a longitudinal study of the US from March to September of 2020. J. Transport Geogr. 93, 103039. https://doi.org/10.1016/j.jtrangeo.2021.103039.

Kraemer, M.U., Yang, C.H., Gutierrez, B., Wu, C.H., Klein, B., Pigott, D.M., Du Plessis, L., Faria, N.R., Li, R., Hanage, W.P., Brownstein, J.S., 2020. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science 368 (6490), 493–497. https://doi.org/10.1126/science.abb4218.

Kwan, M.-P., 2012. The uncertain geographic context problem. Ann. Assoc. Am. Geogr. 102, 958–968. https://doi.org/10.1080/00045608.2012.687349.

Kwan, M.-P., 2021. The stationarity bias in research on the environmental determinants of health. Health Place 70, 102609. https://doi.org/10.1016/j.healthplace.2021.102609.

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

Lai, K.Y., Webster, C., Kumari, S., Sarkar, C., 2020. The nature of cities and the Covid-19 pandemic. Curr. Opin. Environ. Sustain. 46, 27–31. https://doi.org/10.1016/j.cosust.2020.08.008.

Lee, W.D., Qian, M., Schwanen, T., 2021a. The association between socioeconomic status and mobility reductions in the early stage of England's COVID-19 epidemic. Health Place 69, 102563. https://doi.org/10.1016/j.healthplace.2021.102563.

Lee, P.M.Y., Huang, B., Liao, G., Chan, C.K., Tai, L.B., Tsang, C.Y.J., Leung, C.C., Kwan, M.-P., Tse, L.A., 2021b. Changes in physical activity and rest-activity circadian rhythm among Hong Kong community aged population before and during COVID-19. BMC Publ. Health 21, 836. https://doi.org/10.1186/s12889-021-10890-x.

Liu, X., Huang, J., Lai, J., Zhang, J., Senousi, A.M., Zhao, P., 2021. Analysis of urban agglomeration structure through spatial network and mobile phone data. Trans. GIS 25 (4), 1949–1969. https://doi.org/10.1111/tgis.12755.

Marí-Dell'Olmo, M., Gotsens, M., Pasarín, M.I., Rodríguez-Sanz, M., Artazcoz, L., Garcia de Olalla, P., Rius, C., Borrell, C., 2021. Socioeconomic inequalities in COVID-19 in a european urban area: two waves, two patterns. Int. J. Environ. Res. Publ. Health 18 (3), 1256. https://doi.org/10.3390/ijerph18031256.

Mogi, R., Spijker, J., 2021. The influence of social and economic ties to the spread of COVID-19 in Europe. J. Popul. Res. 1–17. https://doi.org/10.1007/s12546-021-09257-1.

Morawska, L., Tang, J.W., Bahnfleth, W., Bluyssen, P.M., Boerstra, A., Buonanno, G., Cao, J., Dancer, S., Floto, A., Franchimon, F., Haworth, C., 2020. How can airborne transmission of COVID-19 indoors be minimised? Environ. Int. 142, 105832. https://doi.org/10.1016/j.envint.2020.105832.

Nakaya, T., Yano, K., 2010. Visualising crime clusters in a space-time cube: an exploratory data-analysis approach using space-time kernel density estimation and scan statistics. Trans. GIS 14 (3), 223–239. https://doi.org/10.1111/j.1467-9671.2010.01194.x.

Nature, 2021. Why indoor spaces are still prime COVID hotspots [online] Available at: https://www.nature.com/articles/d41586-021-00810-9. (Accessed 20 June 2021). Accessed.

Opsahl, T., Colizza, V., Panzarasa, P., Ramasco, J.J., 2008. Prominence and control: the weighted rich-club effect. Phys. Rev. Lett. 101 (16), 168702. https://doi.org/10.1103/PhysRevLett.101.168702.

Oulahen, G., Mortsch, L., Tang, K., Harford, D., 2015. Unequal vulnerability to flood hazards:"ground truthing" a social vulnerability index of five municipalities in Metro Vancouver, Canada. Ann. Assoc. Am. Geogr. 105 (3), 473–495. https://doi.org/10.1080/00045608.2015.1012634.

Raifman, M.A., Raifman, J.R., 2020. Disparities in the population at risk of severe illness from COVID-19 by race/ethnicity and income. Am. J. Prev. Med. 59 (1), 137–139. https://doi.org/10.1016/j.amepre.2020.04.003.

Rizzo, A., Frasca, M., Porfiri, M., 2014. Effect of individual behavior on epidemic spreading in activity-driven networks. Phys. Rev. E 90 (4), 042801. https://doi.org/10.1103/PhysRevE.90.042801.

Sevtsuk, A., Ratti, C., 2010. Does urban mobility have a daily routine? Learning from the aggregate data of mobile networks. J. Urban Technol. 17 (1), 41–60. https://doi.org/10.1080/10630731003597322.

Shim, E., Tariq, A., Chowell, G., 2021. Spatial variability in reproduction number and doubling time across two waves of the COVID-19 pandemic in South Korea, February to July, 2020. J. Global Infect. Dis. 102, 1–9. https://doi.org/10.1016/j.ijid.2020.10.007.

South China Morning Post, 2018. Hong Kong's small flats 'to get even smaller', hitting quality of life [online] Available at: https://www.scmp.com/news/hong-kong/economy/article/2142165/hong-kongs-small-flats-get-even-smaller-hitting-quality-life. (Accessed 20 June 2021). Accessed.

Tate, E., 2013. Uncertainty analysis for a social vulnerability index. Ann. Assoc. Am. Geogr. 103 (3), 526–543. https://doi.org/10.1080/00045608.2012.700616.

The New York Times, 2021. In: 'Coffin Homes' and 'Cages,' Hong Kong Lockdown Exposes Inequality [online] Available at: https://www.nytimes.com/2021/01/26/world/asia/hong-kong-coronavirus-lockdown-inequality.html?_ga=2.185757673.1763379367.1617417488-133072336.1607935695. (Accessed 3 April 2021). Accessed.

Transport Department of Hong Kong, 2020. The Annual Traffic Census, 2019. The Hong Kong Government [online]. Available at: https://www.td.gov.hk/filemanager/en/content_5018/annual%20traffic%20census%202019.pdf. Accessed: 28 March 2021.

World Health Organization, 2020. Coronavirus Disease (COVID-19) Pandemic. Retrieved April 12, 2021, from. https://www.who.int/emergencies/diseases/novel-coronavirus-2019.

Yabe, T., Tsubouchi, K., Fujiwara, N., Wada, T., Sekimoto, Y., Ukkusuri, S.V., 2020. Non-compulsory measures sufficiently reduced human mobility in Tokyo during the COVID-19 epidemic. Sci. Rep. 10, 18053. https://doi.org/10.1038/s41598-020-75033-5.

Yip, T.L., Huang, Y., Liang, C., 2021. Built environment and the metropolitan pandemic: analysis of the COVID-19 spread in Hong Kong. Build. Environ. 188, 107471. https://doi.org/10.1016/j.buildenv.2020.107471.

Zhang, R., Liang, Z., Pang, M., Yang, X., Wu, J., Fang, Y., Ji, H., Qi, X., 2021. Mobility trends and effects on the COVID-19 epidemic—Hong Kong, China. China CDC Weekly 3 (8), 159–161. https://doi.org/10.46234/CCDCW2021.101.