https://doi.org/10.1016/j.scitotenv.2020.144455

Title:

Spatial analysis of the impact of urban geometry and socio-demographic characteristics on COVID-19, a study in Hong Kong

Authors:

Coco Yin Tung Kwok [a], Man Sing Wong [a,b,*], Ka Long Chan [a], Mei-Po Kwan [c,d], Janet Elizabeth Nichol [e], Chun Ho Liu [f], Janet Yuen Ha Wong [g], Abraham Ka Chung Wai [h], Lawrence Wing Chi Chan [i], Yang Xu [a], Hon Li [a], Jianwei Huang [j], Zihan Kan [j]

[a] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China

[b] Research Institute for Sustainable Urban Development, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China

[c] Department of Geography and Resource Management, and Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Hong Kong, China

[d] Department of Human Geography and Spatial Planning, Faculty of Geosciences, Utrecht University, 3584 CB Utrecht, The Netherlands

[e] Department of Geography, School of Global Studies, University of Sussex, UK

[f] Department of Mechanical Engineering, The University of Hong Kong, Hong Kong, China

[g] School of Nursing, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, China

[h] Emergency Medicine Unit, The University of Hong Kong, Hong Kong, China

[i] Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China

[j] Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Hong Kong

[*] Correspondence: Lswong@polyu.edu.hk (M. S. Wong)

Abstract

The World Health Organization considered the wide spread of COVID-19 over the world as a pandemic. There is still a lack of understanding of its origin, transmission, and treatment methods. Understanding the influencing factors of COVID-19 can help mitigate its spread, but little research on the spatial factors has been conducted. Therefore, this study explores the effects of urban geometry and socio-demographic factors on the COVID-19 cases in Hong Kong. For each patient, the places they visited during the incubation period before going to hospital were identified, and matched with corresponding attributes of urban geometry (i.e., building geometry, road network and greenspace) and socio-demographic factors (i.e., demographic, educational, economic, household and housing characteristics) based on the coordinates. The local cases were then compared with the imported cases using stepwise logistic regression, logistic regression with case-control of time, and least absolute shrinkage and selection operator regression to identify factors influencing local disease transmission. Results show that the building geometry, road network and certain socio-economic characteristics are significantly associated with COVID-19 cases. In addition, the results indicate that urban geometry is playing a more important role than the socio-demographic characteristics in affecting COVID-19 incidence. These findings provide a useful reference to the government and the general public as to the spatial vulnerability of COVID-19 transmission and to take appropriate preventive measures in high-risk areas.

Keywords

COVID-19 pandemic, Spatial analysis, Urban geometry, Socio-demographic characteristics

1. Introduction

Coronavirus disease 2019 (COVID-19) is a new emerging infectious virus which is originated in Wuhan, Hubei Province, China (N. Zhu et al., 2020), and has been announced a global pandemic by the World Health Organization (WHO). The earliest known cases were identified in December 2019, and the WHO officially defined COVID-19 as a pandemic on March 11, 2020 (World Health Organization, 2020a). As of June 12, 2020, a cumulative total of 7,410,510 confirmed cases and 418,294 deaths have been reported from 215 countries/regions (World Health Organization, 2020b). Although COVID-19 causes the catastrophic results, the understanding of COVID-19 epidemiology is limited and its mortality is only roughly estimated (Chen et al., 2020; Huang et al., 2020; Ruan et al., 2020).

In recent research, geospatial and spatial-statistical analysis is performed to identify the association between geographic variables and disease incidences. These studies from around the world indicate that COVID-19 cases to be highly correlated with socioeconomic and demographic factors. The socially and economically relevant factors with prominent inter-generational and social characteristics were showing close correlations with COVID-19 in 20 European countries (Mogi and Spijker, 2020). In New York City, the analysis of multi-temporal positive records suggested that occupation is playing a more important role in disease transmission than other factors such as income, race, gender and household size (Almagro and Orane-Hutchinson, 2020). In Sweden, a positive relationship was observed between socio-demographic factors (i.e., gender, income, education level, marital status, immigration status) and the toll of COVID-19 pandemic (Drefahl et al., 2020). Studies performed in Italy and South Korea reported that the ages and the household composition were associated with both the spread and mortality of COVID-19. Mollalo, Vahedi and Rivera (2020) adopted 35 environmental, socioeconomic, topographic and demographic variables to explain the spatial distribution of confirmed cases in the United States, and they concluded that factors such as income, inequality, median household income, race, and the proportion of nursing practitioners were variables with significance. Similarly, Raifman and Raifman (2020) also suggested that race and income are factors that are highly associated with virus exposures in the United States. Khunti et al. (2020) highlighted the association between ethnicity and number of confirmed cases in England and the United States, because the socio-economic conditions are quite different among different racial or ethnic groups in a country. As indicated by Pareek et al. (2020), ethnicity is an important factor that closely related to people's health status, health behavior and social behavior, so the factor should be considered in taking the measures for pandemic control. Therefore, there are pieces of evidence that the socio-economic and demographic characteristics are contributing factors to COVID-19 transmission.

Similar to other infectious diseases, COVID-19 transmission is more frequently observed in urban areas. However, recent studies on COVID-19 did not focus on factors such as urban geometry and urban design that are considered important for the outbreak of other infectious diseases. In England, for example, urbanization promotes the transmission of an infectious disease due to increased contact rates and altered socio-economic conditions (Zhang and Atkinson, 2008). Another study in China suggested that the increased connection between rural and urban areas and the rural-to-urban migration caused by the urbanization could speed up the transmission of infectious disease (Gong et al., 2012). Urban geometry can affect the living environment and should be considered when studying the health conditions of residents, and the factors can also be further divided into sub-factors such as the building geometry, the road network and the distribution of greenspaces, etc. (Johansson, 2006).

A key factor of urban geometry is the Sky View Factor (SVF), which is defined as the ratio of the visible sky with obstructions to that of without obstruction (Oke, 1988), and it is always considered as a typical indicator of urban geometry (Krüger et al., 2011; Yang et al., 2016, 2015). Lai et al. (2013) found that the SVF, which is related to building height and density, would potentially serve as an indicator of the risk of health in an urban area. Some studies stressed the role of air ventilation, which is largely a result of the urban geometry, on disease transmission (Cheng et al., 2011; Gao et al., 2008; Keshavarzian et al., 2020), and poor ventilation was found to be associated with Severe Acute Respiratory Syndrome (SARS) infection (Gao, Li, & Leung, 2009) as well as asthmatic symptoms (Smedje and Norbäck, 2000). It has also been found that the area of outdoor space affects the quality of the indoor environment (Chan and Liu, 2018; Niachou et al., 2008). The former proposed that the neighborhood environment had a direct influence on people's health conditions which is also related to the relationship between the indoor environment and people's health condition. They found that lower density and height of buildings are most beneficial for human health, especially for those with respiratory diseases (e.g., bronchoconstriction, asthma symptoms). Although these studies did not emphasize the infectious disease, they aligned the SARS outbreak in Hong Kong with their findings.

In addition to the building geometry, the road network is another essential factor of city design as it is closely related to traffic flow and the connectivity between different places. It is also found that this factor could facilitate the spread of serious diseases (e.g., malaria and diarrheal pathogens) in developing countries (Coimbra, 1988; Eisenberg et al., 2006) as well as respiratory diseases in developed countries (Vu et al., 2013). Furthermore, Chan and Liu (2018) found that a higher proportion of greenspaces around the buildings are associated with lower levels of air pollution, and such environment is good for people's health, especially for those with respiratory disease. The latest research in China (Y. Zhu et al., 2020) and Western Europe (Ogen, 2020) also found that air pollution was associated with COVID-19 transmission. Therefore, it is important to investigate the role that greenspace plays in COVID-19 transmission (Qu et al., 2020). Based on the findings of these studies, some questions are raised about how urban geometry may affect COVID-19 cases.

With the advancements in Geographical Information System (GIS) technologies, in 1999, Moore and Carpenter (1999) suggested that the relationship between environmental and socio-economic factors and infectious disease displays potential research values. With the use of GIS, we are able to link different types of data spatially, e.g., residential addresses, environmental exposure, building geometry, and demographic information. Indeed, spatial clustering in GIS was adopted for epidemiologic investigations after the outbreak of SARS in Hong Kong in 2003. Using the spatial clustering methods, it was found that SARS was highly clustered disease in Hong Kong as the geospatial clusters were observed in the case locations (Leung et al., 2004), and the urban population is faced with a higher risk (Lai et al., 2004). During COVID-19 pandemic, more than 63 scientific studies that focus on the spatial analysis of COVID-19 have been published, covering spatiotemporal analysis, health and social geography, environmental variables, data mining and web-based mapping (Franch-Pardo et al., 2020).

Spatial variables have long been found to be connected to the infectious disease, and COVID-19 appears to be no exception. However, to date, most of the studies of COVID-19 focused on the socio-demographic factors, while little of them have touched upon the issue of urban geometry. Therefore, it is necessary to analyze the influence of urban geometry, including building configurations, road network and greenspace. This study aims to investigate the importance of spatial context, including urban geometry and socio-demographic factors, in the COVID-19 epidemic in Hong Kong.

2. Study area and data source

2.1 Study area and COVID-19 cases

Hong Kong is selected as the study area for this research (Figure 1). It is a highly urbanized city with a population of more than seven million and it covers an area of 1,111 km^2. Due to its mountainous terrain, the population is squeezed in the densely built and high-rises in urban areas, while these high-rises occupy only approximately 20% of general land area. The first COVID-19 case in Hong Kong was reported on January 23, 2020, which was an imported case from Wuhan, China. Up to June 12, 2020, 1,109 COVID-19 cases were recorded, of which 1,061 cases were discharged, 4 patients were deceased, and the remainder were still hospitalized (Centre for Health Protection (HKSAR), 2020). For this study, the details of COVID-19 cases in Hong Kong were retrieved from the Centre for Health Protection (hereafter government dataset), and Internet source "covid19.vote4.hk - COVID-19 in HK" (hereafter Internet dataset) for the period January 23 to April 30, 2020, covering 1,038 cases.

The government dataset comprised two types of information about the daily cases in Hong Kong. The first type is related to the COVID-19 infection cases in Hong Kong accompanied with the details of individual patients: (1) case number, (2) date of report, (3) date of onset, (4) gender, (5) age, (6) hospital admitted, (7) current status (i.e., hospitalized, discharged, or deceased), (8) citizenship (i.e., Hong Kong or non-Hong Kong resident), (9) case classification (i.e., imported case, close contact of imported case, possibly local case, close contact of possibly local case, local case, or close contact of local case), and (10) confirmation of case (i.e., confirmed or probable). The second type of information contains the residential buildings in which the infected patients have resided or the non-residential buildings with two or more cases in the past 14 days, which is accompanied with detailed such as (1) district, (2) building name, (3) last date of stay of the case(s), and (4) related probable/confirmed cases. Since this dataset provided building names only, the coordinates of the addresses were retrieved by using Google Geocoding API for further spatial analysis. Figure 1 shows the spatial distribution of the patients visited in Hong Kong with their case classification.

Another dataset was retrieved from the Internet dataset which summarizes the reports of the government, Internet, and news media, and it also provides the details of patients' information and high-risk areas. The information regarding the patients was similar to that of the government dataset. The second set of data provides more specific information such as where the patients stayed before hospitalization, the action done (i.e., residence, working, gathering, stay, medical, arrival, departure or transportation) and the coordinates (i.e., latitude and longitude).

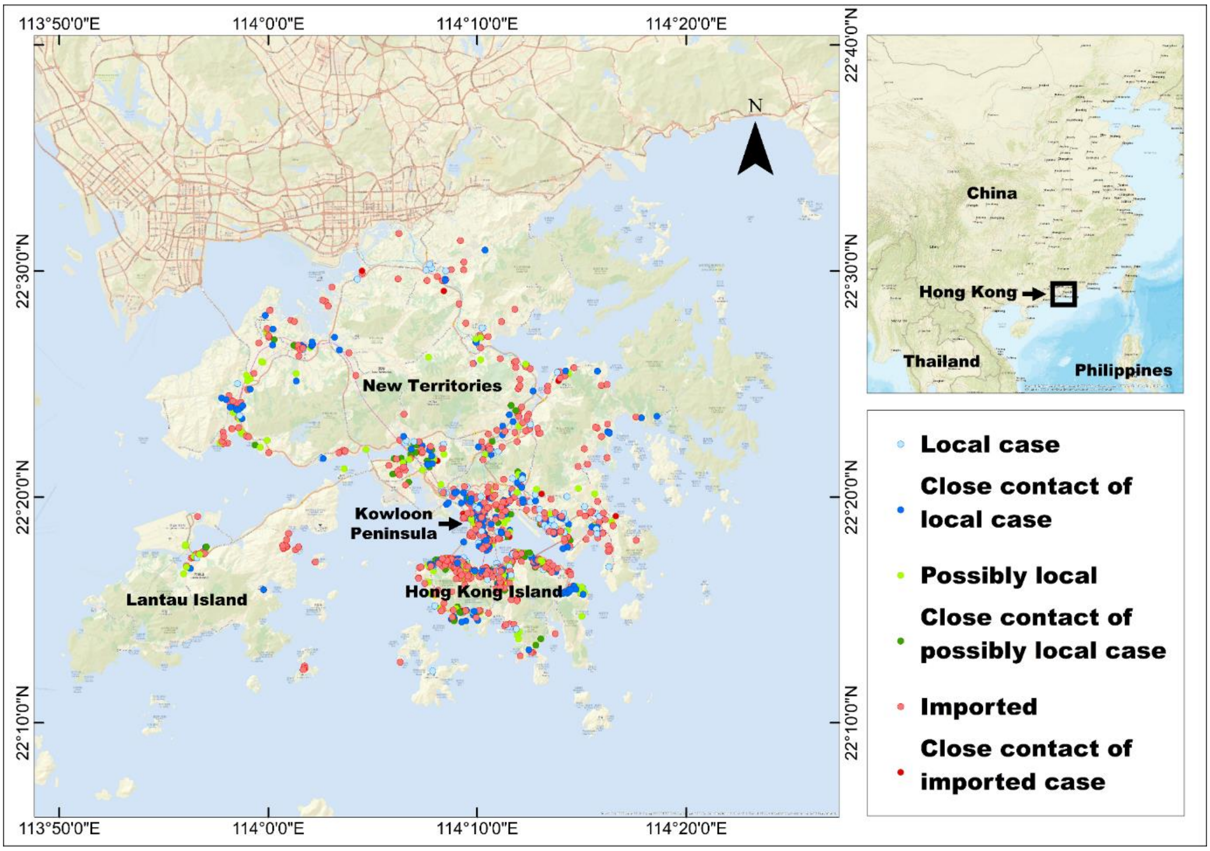

Figure 1 – Study area and the residential spatial distribution of the confirmed cases from government dataset

2.2 Spatial context

(1) Socio-demographic characteristics

The 2016 census data were extracted from the Census and Statistics Department (Census and Statistics Department (HKSAR), 2018) at the Tertiary Planning Unit (TPU) level. The Hong Kong Planning Department uses these regional units for the fine-scale regional planning. The 291 TPUs in Hong Kong were aggregated by the Census and Statistics Department into 154 TPU groups (as shown in Figure 2) to protect personal data privacy of census data. These data include demographic, educational, economic, household and housing characteristics. Specifically, they provide statistics of age, ethnicity, marital status, usual spoken language, reading and writing ability in Chinese and English, educational attainment, economic activity status, monthly income, occupation, industry, working hours per week, place of work, household size, household composition, monthly domestic household income, type of housing, tenure of accommodation and monthly domestic household rent. All the data within the TPUs were transformed from the number of persons into ratio variables, which indicate the percentage of the population within the TPU with certain socio-demographic characteristics. In addition, the total population and the population density of each TPU group were also involved in this study to evaluate the influences of these factors to COVID-19 cases. The detailed information of independent variables adopted in this study is provided in Table A.1 in the appendix. Individual COVID-19 cases were combined with the socio-demographic characteristics data for each TPU by matching the coordinates of places visited by patients in the last 14 days before going to hospital.

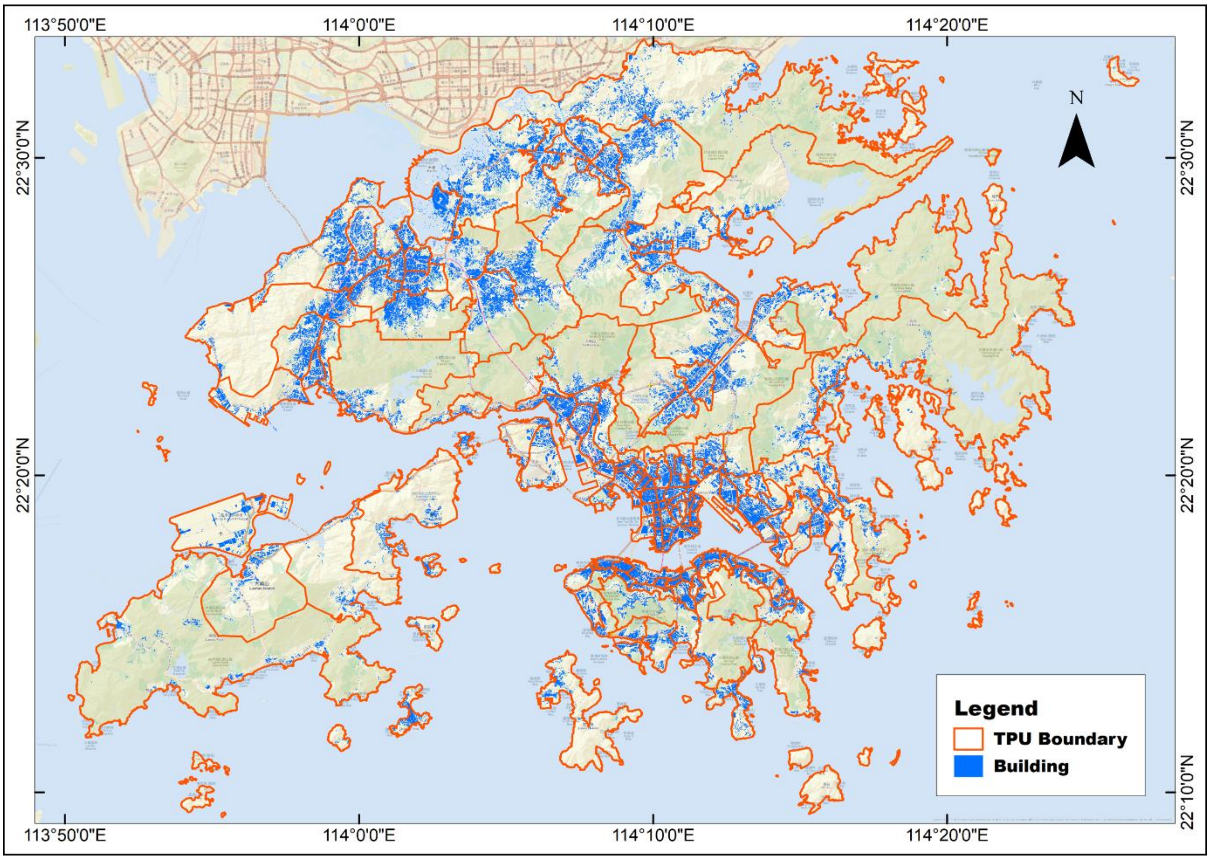

Figure 2 – TPU boundaries, and building data used

(2) Building geometry

In order to understand the role that the urban geometry plays in facilitating the spread of COVID-19, three building-related settings were included in this study (i.e., building height, building density and Sky View Factor (SVF)). Building height was obtained from the Lands Department of Hong Kong in 2019 (Figure 2) based on the 1:1000 scale building polygon that it provides, and then, the data were resampled to 5-meter resolution. Building density was also derived from this building layer by calculating the percentage of building occupation in each 100 m × 100 m area, to measure the crowdedness between buildings (Figure 3). An SVF map at 10-meter resolution (Figure 4) derived by Yang et al. (2015) using airborne LiDAR captured by the Civil Engineering and Development Department of Hong Kong in 2011 was adopted in this study. SVF values range from zero to one, and zero stands for a totally obstructed sky while one for unobstructed sky, respectively. A 500-meter buffer zone was created for individual confirmed cases to study these building-related variables. Within this buffer area, the sum and standard deviation of building height, the building density and SVF were estimated, and these parameters are usually used for representing the urban morphology. The sum of building height, building density and SVF usually represent the urbanization level, where a highly urbanized area always has higher sum values of building height and building density and lower SVF. The standard deviation of these factors represents the variation of the urban morphology and the building geometry. The standard deviation of building density and building height refer to the crowdedness between buildings and the roughness of the buildings in urban morphology. These building related attributes have been extensively utilized in other studies. For example, Hang et al. (2012) evaluated the pollutant dispersion and pedestrian ventilation using different standard deviation of building heights and same average building height to simulate different urban morphologies. Building height, building density and SVF were also used for studying the urban heat island effect caused by urban geometry heterogeneity (Yang and Li, 2015).

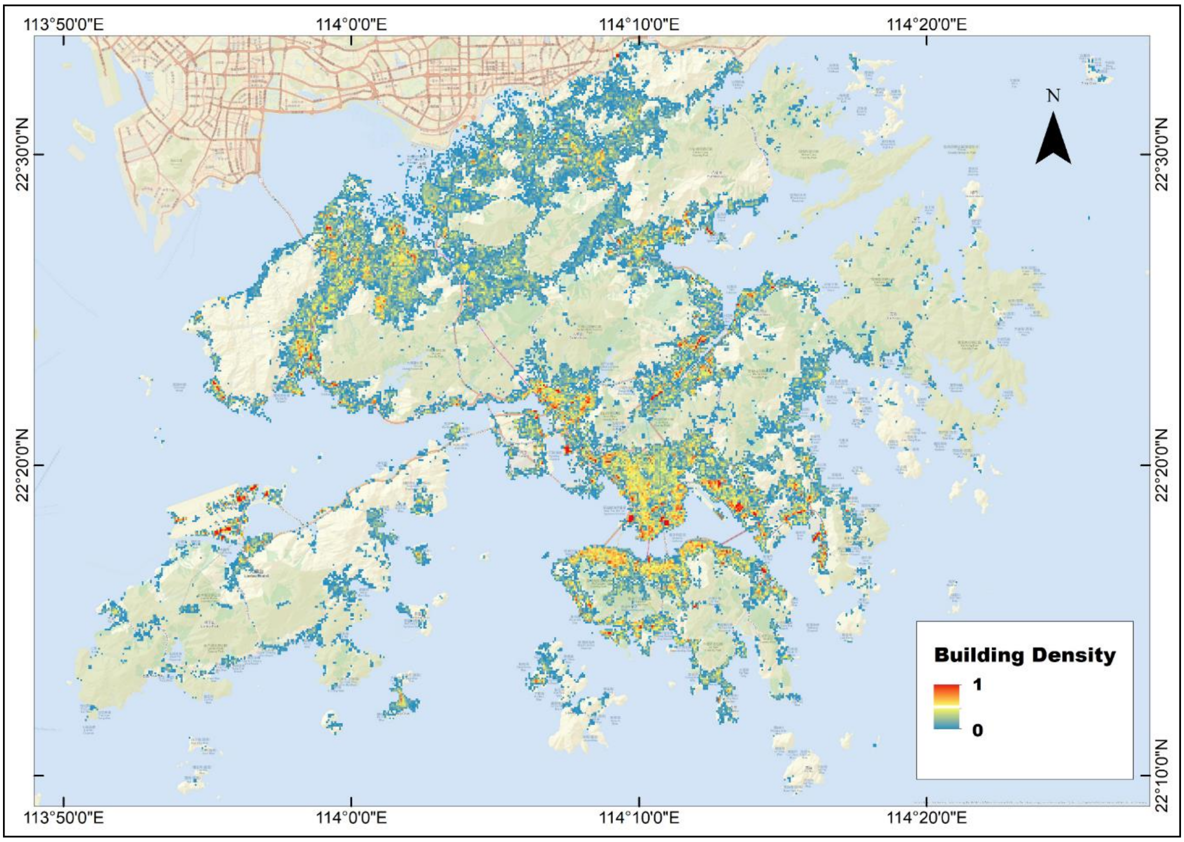

Figure 3 – Map of building density

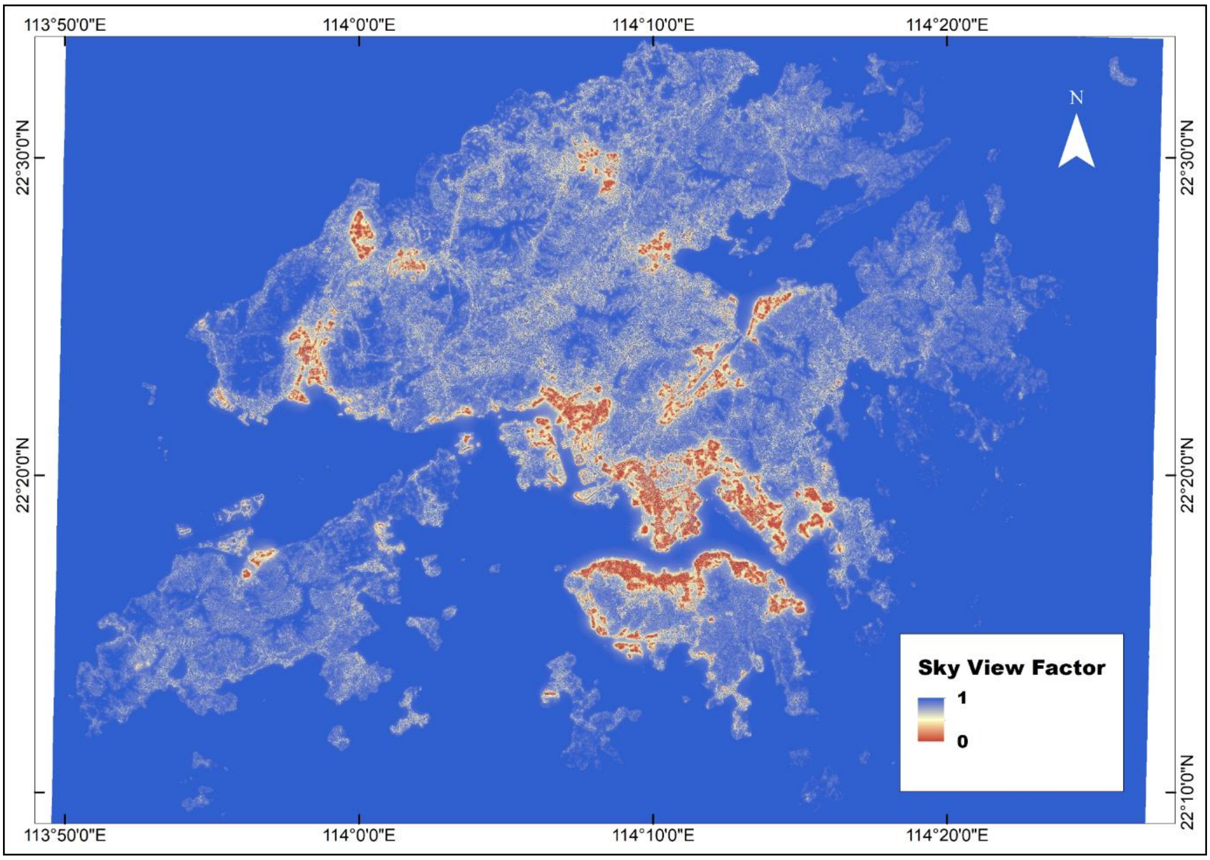

Figure 4 – Map of sky view factor

(3) Road network

In addition to the building geometry, the road networks were also considered, as they are related to the vehicular and pedestrian flows and play an important role in the urban design and planning (Penn et al., 1998). The pedestrian road networks within a 500 m walkable buffer zone of each case were collected with the tool "OSMnx" (Boeing, 2017), road networks can be collected and analyzed easily based on the OpenStreetMap data using graph theory. The characteristics of the road network and the walkable connectivity were considered based on information of nodes, streets, and connectivity in this study. The list of all the parameters adopted is shown in Table A.1.

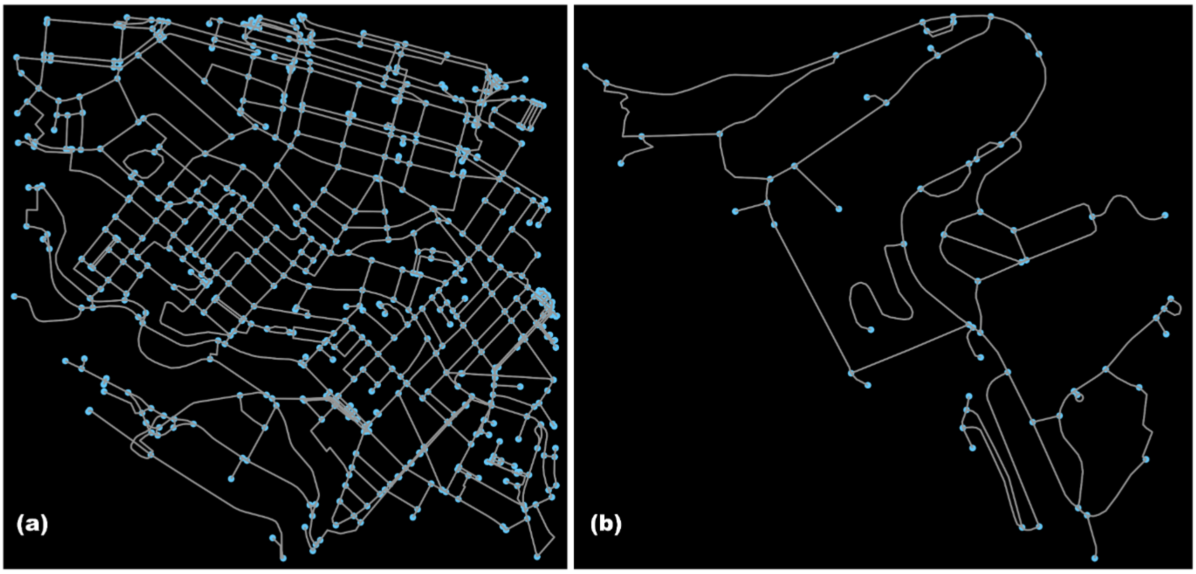

Figure 5 – Examples of road network collected from "OSMnx" (a) urban area; (b) rural area

(4) Greenspace exposure

The neighborhood green space within a 500-meter-radius buffer zone of the residential address of each case was measured using the Normalized Difference Vegetation Index (NDVI) image derived from a Satellite Pour l'Observation de la Terre (SPOT) 7 image captured on February 29, 2016. The NDVI is a normalized ratio of the infrared and red bands, ranging from -1 (no vegetation) to 1 (dense vegetation) (Goward et al., 1991) (Figure 6). The sum and standard deviation of the NDVI within a 500-meter buffer area of each case were retrieved, to represent the total greenspace exposure and the variation of greenspace.

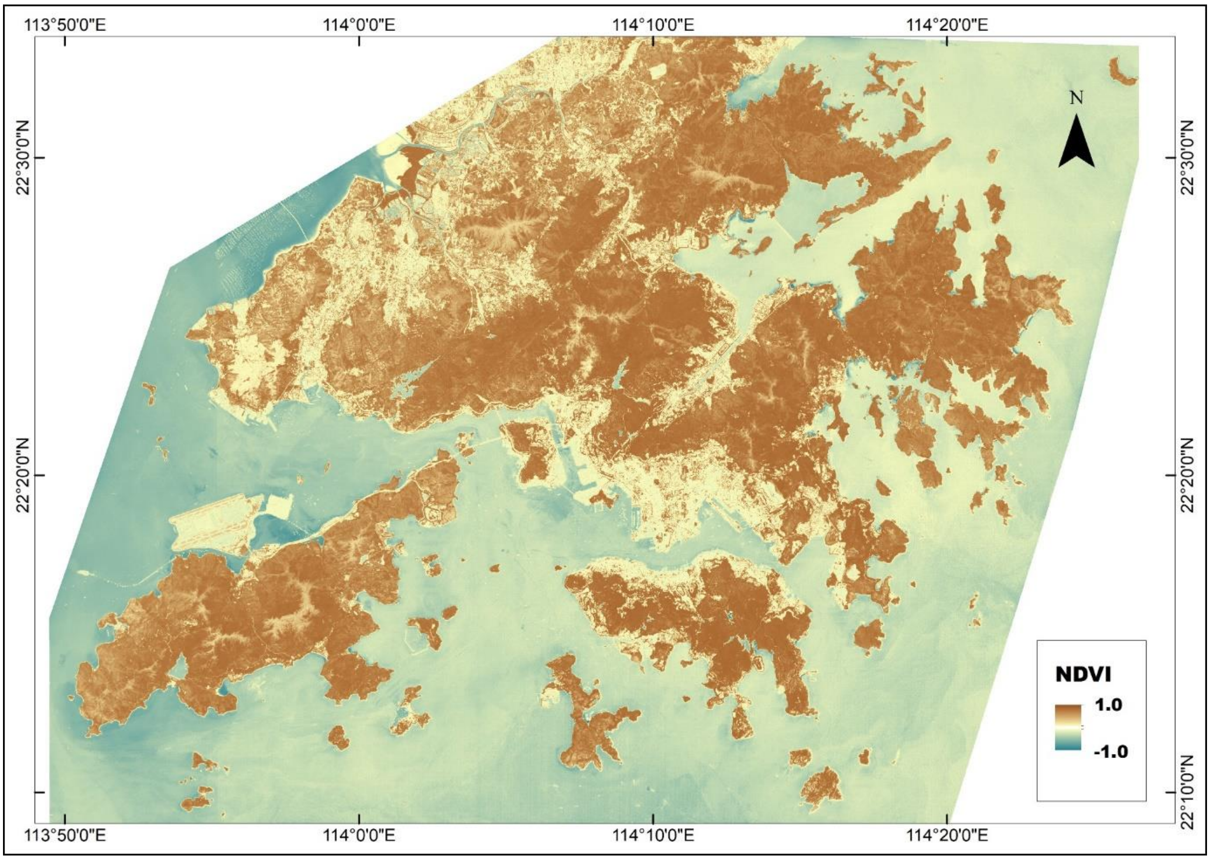

Figure 6 – NDVI map for the calculation of greenspace exposure

3. Methodology

To further investigate the influence of spatial context on COVID-19 transmission, the stepwise logistic regression, the logistic regression with case-control of time, and the least absolute shrinkage and selection operator (Lasso) regression were performed to identify the significant factors. The dependent variable for the regression was the class of cases, that is imported cases or local cases. Local confirmed cases of COVID-19 (including local, possibly local, close contact of local, and close contact of possibly local cases) in Hong Kong were selected as target cases, and non-local cases (imported cases) were selected as the controls, since the non-local cases were not subject to the local factors when they get infected. The close contact with the imported cases was not considered in this study, as they were dependent on both local factors and imported patients.

A total of 125 independent variables were used in this study, as shown in Table A.1 included in the appendix, that is, building geometry, road network, greenspace and socio-demographic characteristics, respectively. These variables were extracted from the government and Internet datasets based on the location list. Hospitals, ports of entry and quarantine facilities were excluded from the list because the patients are infected with COVID-19 before they visit these sites. Information such as neighborhood greenspace, building geometry, road networks, and the census information of individual location retrieved from the government dataset and the Internet were normalized within a range of 0 to 1, to determine the critical risk factors for COVID-19 cases. Spearman's correlation analysis was conducted to evaluate the relationship between variables, and the variables were removed if the correlation value was higher than 0.8 so that the variables with high correlation can be excluded.

3.1 Evaluation of the association between spatial context and the COVID-19 infection

Logistic regression was performed to describe the relationships between independent variables and dependent variable (Kleinbaum and Klein, 2002), while the stepwise approach is adopted to estimate the parameters to be included in the model (Steyerberg et al., 1999). However, there are several limitations if logistic regression is used. First, the date of onset is a crucial factor for understanding the transmission of infectious disease, especially for those diseases which are contagious before the onset

(Fraser et al., 2004). Therefore, the dataset was further analyzed using the conditional logistic regression to confirm the time of case confirmation.

To tackle the problem of the unequal number of cases and control that is commonly occurred in the epidemiological studies, as well as to control the time-related variables, logistic regression with case-control analysis was performed to account for the 1-n matched design. The imported cases were considered as controls and the local cases were considered as the target cases for the analysis in the case-control study. Since the ratio of local cases to imported cases was approximately 1/2 in Hong Kong, two imported case records were matched with each local case record by the closest dates of confirmation for case-control comparison. In addition, in order to conduct temporal evaluation of the influence of the urban geometry and socio-demographic characteristics on COVID-19 local cases, the difference in the confirmation date was set at most three days for each pair of case-control group so that this method evaluated the changes of COVID-19 cases over time among the groups. The commercial software SPSS was adopted to perform the stepwise logistic regression and logistic regression with case-control of time in this study.

To avoid the multicollinearity problem caused by a large number of highly correlated parameters (Shen and Gao, 2008), logistic regression with Lasso regularization was also performed in the current study. Tibshirani (1996) developed Lasso regularization to minimize the sum of squared residuals by applying a penalization parameter to shrink smaller coefficients toward zero, leaving only the most predictive variables in the model:

$$\hat{\beta}^{lasso} = \min(\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}\left|\beta_j\right|) \quad (1)$$

where y_i denotes the dependent variable for the i^{th} data, $x_i = \left(x_{i1}, x_{i2}, \dots, x_{ip}\right)^T$ denotes the predictor variables for the i^{th} data, β_j denotes the coefficient of the regression model for the j^{th} dependent variable, p denotes the number of independent variables, and λ denotes the penalization parameter which is determined by 10-fold cross-validation in this study. This method is useful in identifying the specific factors that were most associated with the confirmed case by eliminating those unassociated variables, so that the predictive performance can be improved (Tibshirani, 2011, 1996). In this study, the logistic regression with Lasso regularization was conducted in R software with “glmnet” package (Friedman et al., 2010; Simon et al., 2011).

4. Results

From January 23, 2020, to April 30, 2020, there were 1,038 COVID-19 cases confirmed in Hong Kong, including 616 imported cases, 25 cases that have been in close contacts with imported cases, 67 local cases, 165 cases that have been in close connection with local cases, 103 likely local cases and 62 cases that have been in close contacts with the likely local cases. The temporal distribution of confirmed cases is shown in Figure 7. The age of patients ranged from 40 days to 96 years, while 559 patients were males and 479 patients were females. Since the cases that have been in close contact with the imported cases were excluded, 1,013 cases were analyzed. Based on the selection criteria as described in the methodology, 1,893 records retrieved from the government dataset and 1,880 records were retrieved from the Internet dataset for logistic regression and Lasso regression. After considering the three-day interval for case-control study, only 1,053 and 1,005 records were retrieved from the government and Internet datasets respectively. The reason why there are more locations than COVID-19 cases is that the patients have been to a number of locations during their incubation period as reported. Some cases reported several locations, and some reported one or none, especially for the imported cases which were diagnosed during entry or the home confinement period. For each reported location, 125 variables were input to the three proposed models after the highly correlated variables (correlation higher than 0.8) were filtered, including six building geometry variables, eight road network variables, two greenspace variables and 109 socio-demographic characteristics.

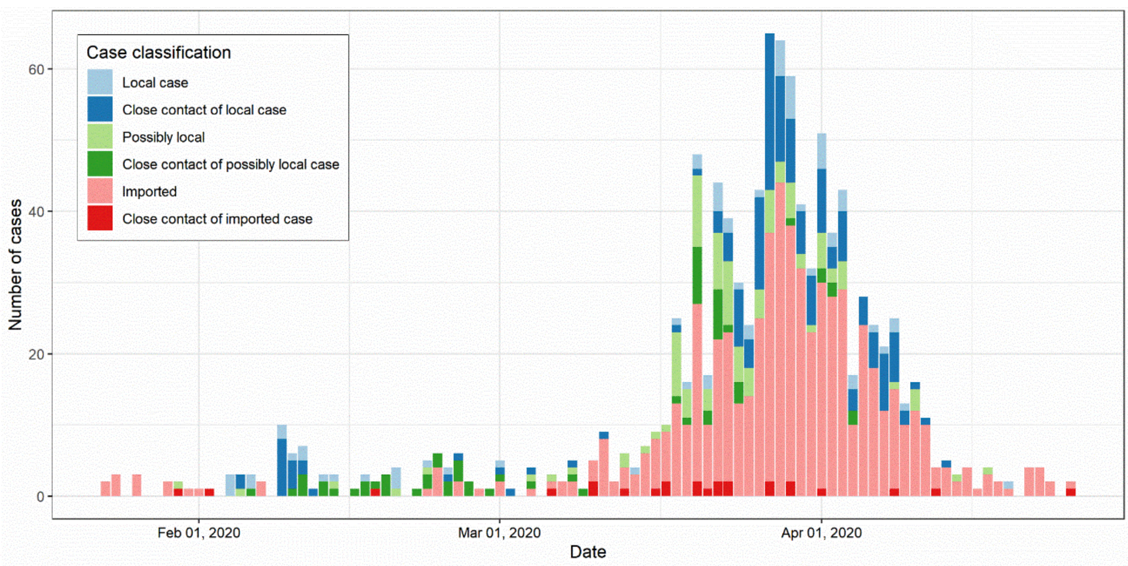

Figure 7 – Temporal distribution of the case class of the confirmed cases

Table 1 presents the coefficients of spatial factors with significance obtained from logistic regression, case-control and Lasso regression for both government and Internet datasets. The insignificant variables were not presented because there are a large number of variables. The independent variables used in logistic regression and case-control analysis were tested using the Chi-square test, and Table 1 shows the variables with p-value smaller than 0.05. For Lasso regression, the coefficients of selected variables from 10-fold cross-validation are listed in Table 1 as well. Among the 125 independent variables, 29 variables were found to have a significant relationship with COVID-19 cases in Hong Kong, in either one of the models or datasets. From the table, it is found that 13 variables displayed a positive relationship, and 16 had a negative relationship with the confirmed cases. Among the significant factors, five variables are associated with the building geometry, one with the road network, two with the demographic characteristics, one with the educational characteristics, 14 with the economic characteristics, one with the household characteristics and five with the housing characteristics. However, no significant relationship was found in its relationship with greenspace exposure.

Table 1 – Coefficient of significant spatial variables from logistic regression analysis, case-control analysis and Lasso regression analysis. The bold text indicates the significant variables from at least three out of six models.

Spatial variable			Government dataset			Internet dataset		
Main category	Sub-category	Variable	Logistic regression	Case-control	Lasso regression	Logistic regression	Case-control	Lasso regression
Urban geometry	Building geometry	**Building height (sum)**	**2.132****	**2.710****	**1.159**	**3.210****	**2.240****	**0.951**
		Building height (standard deviation)	-	-	-	-2.302**	-	-
		Building density (sum)	-	-	-	-	-	0.273
		Building density (standard deviation)	**-1.590****	**-1.797****	**-0.030**	**-1.536****	**-1.924****	**-0.140**
		Sky view factor (sum)	-	-	-0.419	-	-	-0.234
	Road network	**Street length (average)**	**-2.304****	**-**	**-0.675**	**-2.870****	**-**	**-0.226**
Socio-demographic characteristics	Demographic characteristics	Population density	0.849**	-	0.213	-	-	-
		Age group: 65+ (male)	1.007**	-	0.290	-	-	-
	Educational characteristics	Highest educational attainment: Sub-degree course	-	-	-	1.316**	-	0.100
	Economic characteristics	Economic status: Others	-	-	-	-	-	0.298
		Occupation: Professionals	-	-	-	-2.543**	-1.669**	-
		Occupation: Service and sales workers	**-**	**-**	**0.093**	**1.278***	**-**	**0.325**
		Occupation: Craft and related workers	-	-	-	-3.459**	-	-
		Occupation: Skilled agricultural and fishery workers; and occupations not classifiable	**-2.123****	**-3.475****	**-0.719**	**-**	**-**	**-**
		Industry: Accommodation and food services	-	-	-	-	-	0.415
		Industry: Manufacturing	-	-	-	-	-	-0.049

		Industry: Public administration, education, human health and social work activities	-	-	-	-	-	-0.183
		Working location: Another district on Hong Kong Island#	**-**	**-**	**0.153**	**0.813****	**1.058****	**0.638**
		Working location: Outside Hong Kong	-	-	-	-1.116**	-	-
		Weekly working hours: 18-34	-	-	-	-	-	0.108
		Weekly working hours: 65 and over	-	1.458*	-	-	-	-
		Median monthly income from main employment (male)	-	-1.723**	-	-	-	-
		Median monthly income from main employment (female)	-	-	-	-	-	-0.073
	Household characteristics	Median monthly domestic household income	-	-	-	-	-	-0.732
	Housing characteristics	Tenure of accommodation: Owner-occupier (with mortgage and loan)	-	-	-0.018	-	-	-
		Tenure of accommodation: Owner-occupier (without mortgage and loan)	-	-	-	-1.421**	-	-
		Tenure of accommodation: Sole tenant	-	-	-	-	-	0.309
		Tenure of accommodation: Co-tenant	-0.642**	-	-	-0.705*	-	-
		Tenure of accommodation: Provided by employer	-	-	-	-1.229*	-	-

** Significant result with p-value < 0.01

* Significant result with p-value < 0.05

"Working location: Another district on Hong Kong Island" means the number of persons working on Hong Kong Island, excluding the persons living and working in the same district on Hong Kong Island.

The results from the three models using the two datasets are not identical, we selected those factors which were significant in at least three models for the investigation, and there are six factors in total (Table 2). Three factors are associated with urban geometry and three are associated with the socio-demographic characteristics. These factors were the sum of building height (positive in six models), the standard deviation of building density (negative in six models), average street length (negative in four models), working location in another district on Hong Kong Island (i.e., the number of persons working on Hong Kong Island, excluding the persons living and working in the same district on Hong Kong Island) (positive in five models), service and sales workers (positive in three models), skilled agricultural, fishery workers and occupations not classifiable (negative in three models).

Table 2 – Summary of important spatial variables from at least three models

Spatial variable			Number of models indicated as significant factors	Sign of the coefficient in the model
Main category	Sub-category	Variable		
Urban geometry	Building geometry	Building height (sum)	6	(+)
		Building density (standard deviation)	6	(-)
	Road network	Street length (average)	4	(-)
Socio-demographic characteristics	Economic characteristics	Working location: Another district on Hong Kong Island	4	(+)
		Occupation: Service and sales workers	3	(+)
		Occupation: Skilled agricultural and fishery workers; and occupations not classifiable	3	(-)

Table 3 generalizes the absolute contribution of the spatial factors (i.e., sum of the absolute coefficient of the factors listed in Table 2) to each of the models. Since the values were normalized through pre-processing phases, these weightings can be directly compared. Comparison of the coefficients of urban geometry with the socio-demographic characteristics shows that the weighting of urban geometry is higher than that of socio-demographic characteristics in all models. By analyzing the ratio of the important factors between urban geometry and socio-demographic characteristics as shown in brackets of Table 3, it is found that the number of variables related to the urban geometry is usually greater than or equal to the socio-demographic characteristics. These two findings indicate the importance of urban geometry to COVID-19 cases in Hong Kong.

Table 3 – Weighting of the important spatial variables to COVID-19 cases based on the main category and the number in brackets indicates the important factors from the model

Spatial variable	Government dataset			Internet dataset		
	Logistic regression	Case-control	Lasso regression	Logistic regression	Case-control	Lasso regression
Urban geometry	6.026 (3)	4.507 (2)	1.863 (3)	7.616 (3)	4.164 (2)	1.317 (3)
Socio-demographic characteristics	2.123 (1)	3.475 (1)	0.965 (3)	2.091 (2)	1.058 (1)	0.963 (2)

5. Discussion

This study compared the local with the imported COVID-19 cases in Hong Kong with the use of logistic, case-control and Lasso regressions to identify the relevant spatial factors in the transmission and spread of the disease. A total of 125 variables were analyzed with six of them related to the building geometry, eight related to the road networks, two related to the greenspace and 109 related to the socio-demographic characteristics. Six important factors were determined through six models, to explain the relationship between urban geometry, socio-demographic characteristics and the incidence of COVID-19.

Of the socio-demographic characteristics investigated, those found to be associated with COVID-19 patients were related to the working locations and occupations, which are consistent with the findings of other studies (Almagro and Orane-Hutchinson, 2020). Positively related factors included the working location on Hong Kong Island and the occupation as service and sales workers. Hong Kong Island is the central business district of Hong Kong, which occupies 7% of the land area, while but 23% of the working population resides on the island. High population density has been discussed in several studies, as one of the factors leading to the transmission or mortality of COVID-19 (Rocklöv & Sjödin, 2020; Wu et al., 2020). The densely populated Hong Kong Island provides many chances of social contacts among the large working population. In addition, the workers engaged in the service and sales industry have a greater chance of getting close the contact with the customers and clients, so they are faced with greater risks due to the nature of their jobs. This finding aligned with findings from the first outbreak of the COVID-19 in Wuhan, China, and those earliest confirmed cases were salesmen or saleswomen at Huanan Seafood Wholesale Market (Rothan and Byrareddy, 2020). Based on the experience from Singapore, Koh (2020) also suggested that occupational exposure are the main reasons for earliest confirmed cases engaged in the tourism, retail and transportation industries. The result of this study further supports the findings of these previous studies, indicating high exposure risk of certain occupations. In contrast, the occupation of the group "skilled agricultural, fishery workers and occupations not classifiable" was negatively correlated with the infection. These workers are either working outdoors or tend to work at one location away from urban areas (e.g., cultivated field, mariculture raft, fishing boat or home) leading to lower contact with others.

Some of the other studies highlighted the importance of the socio-demographic characteristics on COVID-19 cases, especially the population density on the COVID-19 case number and the spread of the disease. For example, direct relationship was found between the COVID-19 outbreak and the population density in Iran (Ahmadi et al., 2020). Similar finding from the study in Japan suggested the positive correlations between the morbidity and mortality rates and population density (Kodera et al., 2020). The study in Turkey emphasized the importance of the population density and wind that these two factors can explain 94% of the variance in the spread of COVID-19 in the country (Coşkun et al., 2021). However, most of the studies have not yet considered the urban geometry as a factor in their analysis. In this study, the population density was found to be less important than the urban geometry to COVID-19 cases. From Table 1, population density is the significant factor in two out of six models only while the sum of building height and standard deviation of building density are involved in all six models. In logistic regression and Lasso regression using government dataset, the coefficients of population density are 0.849 and 0.213 respectively, while the coefficients of the sum of building height are 2.132 and 1.159, in which the standard deviation of building density are -1.590 and -0.030, and that of average street length are -2.304 and -0.675 respectively. When comparing the coefficients of the variables, it is clearly observed that the contribution of urban geometry characteristics is higher than the population density in the models.

To our understanding, there was no research on the relationship between COVID-19 and urban geometry in the past. The results of the study reveal the importance of urban geometry to COVID-19. The most important spatial factors found in this study were the building height and the building density, and they were two variables with significance in all six models. There was a positive association between COVID-19 cases and the sum of building height, and a negative association with the standard deviation of building density. Building height and building density are key components of urban geometry and surface roughness, and these greatly determine the magnitude of wind ventilation (Kubota et al., 2008; Rafailidis, 1997; Wong et al., 2010). The total building height evaluates the level of urbanization level in an area, as high buildings usually exist in urban areas where there are densely built high-rises. In Hong Kong, there are 20 floors to 50 floors in most urban buildings (Wong et al., 2010), and high-rise and dense buildings are always considered as factors that are not good for wind ventilation, and in such an environment, people's health condition will be adversely affected (Wargocki, 2013). Building density is an indicator of how dense of the buildings in an area, and the wind ventilation is always poor in the high building density area since the wind is obstructed by the buildings (Yang et al., 2019). The standard deviation of building density can represent the building distribution of the area, and the higher building density (standard deviation) means more diversified the building blockage. More diversified building blockage improves ventilation because of the enhanced turbulence generation by rough surfaces. A typical case of the building environment affecting the spread of infectious disease was happened during SARS outbreak in 2003, where more than 300 people who lived in Amoy Gardens housing complex in Hong Kong were infected. The consensus document of the World Health Organization reported the large number of infection in building Block E that the dry U-traps in bathroom allowed the contaminated sewage droplets entering households, and the increasing number of virus in the sewer system was aerosolized in bathroom (World Health Organization, 2003) and contaminated droplets were transported through the running exhaust fan to air shaft and thus to the upper part of the building (World Health Organization Regional Office for the Western Pacific, 2003). Since some of the patients did not have person-to-person contacts with others, Yu et al. (2004) conducted some epidemiologic analysis, experimental studies and airflow simulations to further examinate and confirm the airborne spread of SARS between building blocks. Since the outbreak of SARS and avian flu in 2003, researchers start investigating the relationship between urban environments and the infectious diseases, and how different urban settings can affect health of inhabitants and the spread of diseases (Wolf, 2016). This study further emplasizes the relationship between building geometry and COVID-19 cases.

The average street length was significantly related to urban geometry, as four models suggested the negative correlation of this factor with the infection risk within a 500m walkable zone. Longer lengths generally represented by main roads and shorter lengths represented by short streets and lanes. As shown in Figure 5a, the pedestrian networks in Hong Kong urban areas usually consist of several main roads and a number of short streets connecting with the main roads. Long average street length is the result of low connectivity where only main roads exist in the target zone (Figure 5b) and there are fewer connections between roads. The walking mobility is thus restricted by limited connectivity and this also reduces the social contact. Increasing social distancing has been demonstrated as a mean to reduce the spread of COVID-19, and the low walking mobility from low connectivity is able to achieve the objective in reducing the infection risk, probably through increased social distancing.

In addition to the analysis of individual factors, a major conclusion of this study is that urban geometry is more important than socio-demographic characteristics in COVID-19 risk. Most other studies of the spatial-statistical factors related to COVID-19 cases focused on socio-demographic characteristics, (e.g., Drefahl et al., 2020; Mogi & Spijker, 2020; Raifman & Raifman, 2020) while the urban geometry is rarely considered. In this study, although the number of independent variables for urban geometry was fewer than socio-demographic characteristics, the results indicate that urban geometry is more important than socio-demography in affecting the COVID-19 cases in Hong Kong. Although this significant association has not been mentioned in the literature, urban geometry has been found influential to other diseases. For example, building properties were found correlated with tuberculosis (Lai et al., 2013), respiratory conditions (McCarthy et al., 1985), visual and acoustic comfort (Chan and Liu, 2018), thermal comfort (Ali-Toudert and Mayer, 2006) and excess mortality (Wong et al., 2017). The results of this study confirm the association between urban geometry and disease in the case of COVID-19. The findings on the importance of the urban geometry in this study can be further extended to other infectious diseases, e.g. influenza, tuberculosis and dengue fever. Since there are only a few studies evaluating the association between infectious diseases and urban geometry, this study could further be extended once the data are available.

Limitations of the study include the use of the imported cases as a control group for comparison with the local cases, as it is based on the assumption that the imported cases were not exposed in the high-risk areas when they became infected. The reason for not selecting healthy people as the control group was to prevent bias caused by the differences between the control group and target group, as more than seven million people did not catch the disease. Regarding the data used, all the urban geometry and socio-demographic factors considered were collected at a fixed time, as the spatially and temporally dynamic data were unavailable. In addition, the socio-demographic data have been retrieved from the census at TPU, rather than based on the characteristics of an individual patient. The method used in this study displays great applicability in determining the influence of the social environment on the individuals, rather than the individual personal characteristics (Kosatsky et al., 2012). If the personal characteristics of individual patients could be obtained, an individual vulnerability could be identified using a similar approach. Although this study could not exhaust all the potential factors, it provides a useful approach to the identification and highlights the importance of urban geometry to COVID-19 cases.

6. Conclusions

In this study, a combination of logistic, case-control and Lasso regressions was performed to evaluate the importance of urban geometry and socio-demographic factors in the transmission and spread of COVID-19 cases in Hong Kong. The main factors contributing to increased risk were building height, working in another district on Hong Kong Island, and service and sales occupations. Low-risk factors included districts with large variation in building density, low walkability and occupation of skilled agricultural and fishery workers, and occupations not classifiable. The results suggest the important contribution of urban design and geometry, including building geometry and road network settings, to risk from COVID-19 when compared with socio-demographic characteristics. This result can provide insight for citizens to understand and avoid risk, and for government to establish planning and design policies to minimize disease transmission in the short-term and better urban planning in the long-term.

Acknowledgements

The authors would like to thank the Centre for Health Protection of HKSAR and the internet source "covid19.vote4.hk - COVID-19 in HK" for the COVID-19 patients' data, the Census and Statistics Department of HKSAR for the census data, the Lands Department of HKSAR for the building data, and the Civil Engineering and Development Department of HKSAR for the airborne LiDAR data.

Funding

This research is partially supported by grants (Grant no. 15602619 and C7064-18GF) from the Research Grants Council of Hong Kong SAR, China; and the Research Institute for Sustainable Urban Development (Grant no. 1-BBWD), the Hong Kong Polytechnic University, Hong Kong, China. Mei-Po Kwan was supported by a grant from the Research Committee on Research Sustainability of Major RGC Funding Schemes of the Chinese University of Hong Kong, Hong Kong, China.

Appendix

Table A.1 – Variables used in this study

Main category	Sub-category	Variable
Urban geometry	Building geometry	1. Building height (Sum/Standard deviation) 2. Building density (Sum/Standard deviation) 3. Sky view factor (Sum/Standard deviation)
	Road network	1. Number of nodes in network 2. Number of edges in network 3. Average node degree 4. Intersection count 5. Average streets per node 6. Counts of streets per node 7. Total edge length 8. Average edge length 9. Total street length 10. Average street length 11. Count of street segments 12. Average circuity 13. Self-loop proportion 14. Mean average neighbourhood degree 15. Mean average weighted neighbourhood degree 16. Average degree centrality 17. Average weighted clustering coefficient 18. Average betweenness centrality
	Greenspace	1. Normalized difference vegetation index (Sum/Standard deviation)
Socio-demographic characteristics	Demographic characteristics	1. Total number of populations 2. Population density 3. Age (i) 0 – 19 (Male/Female/Both sex) (ii) 10 – 64 (Male/Female/Both sex) (iii) 65 + (Male/Female/Both sex) 4. Median age (Male/Female/Both sex) 5. Ethnicity (i) Chinese (ii) Filipino (iii) Indonesian (iv) White (v) Others 6. Marital Status (i) Never married (ii) Married (iii) Widowed (iv) Divorced (v) Separated 7. Usual spoken language

		(i) Cantonese (ii) Putonghua (iii) Other Chinese dialects (iv) English (v) Other languages 8. Whether able to read Chinese (i) Able to read (ii) Not able to read 9. Whether able to read English (i) Able to read (ii) Not able to read 10. Whether able to write Chinese (i) Able to write (ii) Not able to write 11. Whether able to write English (i) Able to write (ii) Not able to write
	Educational characteristics	1. Educational attainment (highest level attended) (i) No schooling/ Pre-primary (ii) Primary (iii) Lower secondary (iv) Upper secondary (v) Post-secondary: Diploma/ Certificate (vi) Post-secondary: Sub-degree course (vii) Post-secondary: Degree course
	Economic characteristics	1. Economic activity status (i) Employees (ii) Employers (iii) Self-employed (iv) Unpaid family workers (v) Home-makers (vi) Students (vii) Retired (viii) Others 2. Place of Work (i) Work in the same district (ii) Work in another district on Hong Kong Island (iii) Work in another district in Kowloon (iv) Work in another district in New Towns (v) Work in another district in other areas in the New Territories (vi) No fixed place/ Marine (vii) Work at home (viii) Places outside Hong Kong 3. Monthly income from main employment (i) < HK$10,000 (ii) HK$10,000 –HK$19,999

		(iii) HK$20,000 –HK$39,999 (iv) > = HK$ 40,000 4. Median monthly income from main employment (Male/Female/Both sex) 5. Occupation (i) Managers and administrators (ii) Professionals (iii) Associate professionals (iv) Clerical support workers (v) Service and sales workers (vi) Craft and related workers (vii) Plant and machine operators and assemblers (viii) Elementary occupations (ix) Skilled agricultural and fishery workers; and occupations not classifiable 6. Industry (i) Manufacturing (ii) Construction (iii) Import/export, wholesale and retail trades (iv) Transportation, storage, postal and courier services (v) Accommodation and food services (vi) Information and communications (vii) Financing and insurance (viii) Real estate, professional and business services (ix) Public administration, education, human health and social work activities (x) Miscellaneous social and personal services (xi) Others: including “Agriculture; forestry and fishing”; “Mining and quarrying”; “Electricity and gas supply”; “Water supply; sewerage, waste management and remediation activities” and industrial activities unidentifiable or inadequately described 7. Weekly usual hours of work of all employment (i) < 18 (ii) 18 – 34 (iii) 35 – 44 (iv) 45 – 54 (v) 55 – 64 (vi) 65+
	Household characteristics	1. Household size (i) 1 (ii) 2 (iii) 3 (iv) 4 (v) 5

		(vi) 6 + 2. Average domestic household size 3. Household composition (i) Composed of couple (ii) Composed of couple and unmarried children (iii) Composed of lone parent and unmarried children (iv) Composed of couple and at least one of their parents (v) Composed of couple, at least one of their parents and their unmarried children (vi) Composed of other relationship combinations (vii) One-person households (viii) Non-relative households 4. Monthly domestic household income (i) < HK$10,000 (ii) HK$10,000 –HK$19,999 (iii) HK$20,000 –HK$39,999 (iv) HK$40,000 –HK$79,999 (v) > = HK$ 80,000 5. Median monthly domestic household income 6. Median monthly household income of economically active households
	Housing characteristics	1. Type of Housing (i) Public rental housing (ii) Subsidised home ownership housing (iii) Private permanent housing (iv) Non-domestic housing (v) Temporary housing 2. Tenure of Accommodation (i) Owner-occupier – With mortgage or loan (ii) Owner-occupier – Without mortgage and loan (iii) Sole tenant (iv) Co-tenant/Main tenant/Sub-tenant (v) Rent free (vi) Provided by employer 3. Median monthly domestic household rent 4. Median rent to income ratio

References

Ahmadi, M., Sharifi, A., Dorosti, S., Jafarzadeh Ghoushchi, S., Ghanbari, N., 2020. Investigation of effective climatology parameters on COVID-19 outbreak in Iran. Sci. Total Environ. 729, 138705. https://doi.org/10.1016/j.scitotenv.2020.138705

Ali-Toudert, F., Mayer, H., 2006. Numerical study on the effects of aspect ratio and orientation of an urban street canyon on outdoor thermal comfort in hot and dry climate. Build. Environ. 41, 94–108.

Almagro, M., Orane-Hutchinson, A., 2020. The determinants of the differential exposure to COVID-19 in New York City and their evolution over time. Covid Econ. Vetted Real-Time Pap.

Boeing, G., 2017. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. Comput. Environ. Urban Syst. 65, 126–139. https://doi.org/10.1016/j.compenvurbsys.2017.05.004

Census and Statistics Department (HKSAR), 2018. 2016 Population By-Census [WWW Document]. URL https://www.bycensus2016.gov.hk/en/index.html (accessed 6.17.20).

Centre for Health Protection (HKSAR), 2020. Latest situation of cases of COVID-19 (as of 12 June 2020) [WWW Document]. URL https://www.chp.gov.hk/files/pdf/local_situation_covid19_en.pdf (accessed 6.12.20).

Chan, I.Y.S., Liu, A.M.M., 2018. Effects of neighborhood building density, height, greenspace, and cleanliness on indoor environment and health of building occupants. Build. Environ. 145, 213–222. https://doi.org/10.1016/j.buildenv.2018.06.028

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y., Wei, Y., Xia, J., Yu, T., Zhang, X., Zhang, L., 2020. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 395, 507–513. https://doi.org/10.1016/S0140-6736(20)30211-7

Cheng, C.K.C., Lam, K.M., Leung, Y.T.A., Yang, K., Li Danny, H.W., Cheung Sherman, C.P., 2011. Wind-induced natural ventilation of re-entrant bays in a high-rise building. J. Wind Eng. Ind. Aerodyn. 99, 79–90. https://doi.org/10.1016/j.jweia.2010.11.002

Coimbra, C., 1988. Human Factors in the Epidemiology of Malaria in the Brazilian Amazon. Hum.

Organ. 47, 254–260. https://doi.org/10.17730/humo.47.3.j18mn0682146021q

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

Drefahl, S., Wallace, M., Mussino, E., Aradhya, S., Kolk, M., Brandén, M., Malmberg, B., Andersson, G., 2020. Socio-demographic risk factors of COVID-19 deaths in Sweden: A nationwide register study. Stockholm Research Reports in Demography. https://doi.org/10.17045/STHLMUNI.12420347.V2

Eisenberg, J.N.S., Cevallos, W., Ponce, K., Levy, K., Bates, S.J., Scott, J.C., Hubbard, A., Vieira, N., Endara, P., Espinel, M., Trueba, G., Riley, L.W., Trostle, J., 2006. Environmental change and infectious disease: How new roads affect the transmission of diarrheal pathogens in rural Ecuador. Proc. Natl. Acad. Sci. U. S. A. 103, 19460–19465. https://doi.org/10.1073/pnas.0609431104

Franch-Pardo, I., Napoletano, B.M., Rosete-Verges, F., Billa, L., 2020. Spatial analysis and GIS in the study of COVID-19. A review. Sci. Total Environ. 739, 140033. https://doi.org/10.1016/j.scitotenv.2020.140033

Fraser, C., Riley, S., Anderson, R.M., Ferguson, N.M., 2004. Factors that make an infectious disease outbreak controllable. Proc. Natl. Acad. Sci. U. S. A. 101, 6146–6151. https://doi.org/10.1073/pnas.0307506101

Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33, 1–22. https://doi.org/10.18637/jss.v033.i01

Gao, N.P., Niu, J.L., Perino, M., Heiselberg, P., 2008. The airborne transmission of infection between flats in high-rise residential buildings: Tracer gas simulation. Build. Environ. 43, 1805–1817. https://doi.org/10.1016/j.buildenv.2007.10.023

Gao, X., Li, Y., Leung, G.M., 2009. Ventilation Control of Indoor Transmission of Airborne Diseases in an Urban Community. Indoor Built Environ. 18, 205–218. https://doi.org/10.1177/1420326X09104141

Gong, P., Liang, S., Carlton, E.J., Jiang, Q., Wu, J., Wang, L., Remais, J. V., 2012. Urbanisation and health in China. Lancet 379, 843–852. https://doi.org/10.1016/S0140-6736(11)61878-3

Goward, S.N., Markham, B., Dye, D.G., Dulaney, W., Yang, J., 1991. Normalized difference vegetation index measurements from the advanced very high resolution radiometer. Remote Sens. Environ. 35, 257–277. https://doi.org/10.1016/0034-4257(91)90017-Z

Hang, J., Li, Y., Sandberg, M., Buccolieri, R., Di Sabatino, S., 2012. The influence of building height variability on pollutant dispersion and pedestrian ventilation in idealized high-rise urban areas. Build. Environ. 56, 346–360. https://doi.org/10.1016/j.buildenv.2012.03.023

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., Xiao, Y., Gao, H., Guo, L., Xie, J., Wang, G., Jiang, R., Gao, Z., Jin, Q., Wang, J., Cao, B., 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395, 497–506. https://doi.org/10.1016/S0140-6736(20)30183-5

Johansson, E., 2006. Influence of urban geometry on outdoor thermal comfort in a hot dry climate: A study in Fez, Morocco. Build. Environ. 41, 1326–1338. https://doi.org/10.1016/j.buildenv.2005.05.022

Keshavarzian, E., Jin, R., Dong, K., Kwok, K.C.S., Zhang, Y., Zhao, M., 2020. Effect of pollutant source location on air pollutant dispersion around a high-rise building. Appl. Math. Model. 81, 582–602. https://doi.org/10.1016/j.apm.2020.01.019

Khunti, K., Singh, A.K., Pareek, M., Hanif, W., 2020. Is ethnicity linked to incidence or outcomes of covid-19? BMJ 369. https://doi.org/10.1136/bmj.m1548

Kleinbaum, D.G., Klein, M., 2002. Logistic Regression, 3rd ed. Springer-Verlag, New York.

Kodera, S., Rashed, E.A., Hirata, A., 2020. Correlation between COVID-19 Morbidity and Mortality Rates in Japan and Local Population Density, Temperature, and Absolute Humidity. Int. J. Environ. Res. Public Health 17, 5477. https://doi.org/10.3390/ijerph17155477

Koh, D., 2020. Occupational risks for COVID-19 infection. Occup. Med. (Lond). 70, 3.

Kosatsky, T., Henderson, S.B., Pollock, S.L., 2012. Shifts in mortality during a hot weather event in Vancouver, British Columbia: rapid assessment with case-only analysis. Am. J. Public Health 102, 2367–71. https://doi.org/10.2105/AJPH.2012.300670

Krüger, E.L., Minella, F.O., Rasia, F., 2011. Impact of urban geometry on outdoor thermal comfort and air quality from field measurements in Curitiba, Brazil. Build. Environ. 46, 621–634. https://doi.org/10.1016/j.buildenv.2010.09.006

Kubota, T., Miura, M., Tominaga, Y., Mochida, A., 2008. Wind tunnel tests on the relationship between building density and pedestrian-level wind velocity: Development of guidelines for realizing acceptable wind environment in residential neighborhoods. Build. Environ. 43, 1699–1708. https://doi.org/10.1016/j.buildenv.2007.10.015

Lai, P.-C., Low, C.-T., Tse, W.-S.C., Tsui, C.-K., Lee, H., Hui, P.-K., 2013. Risk of tuberculosis in high-rise and high density dwellings: An exploratory spatial analysis. Environ. Pollut. 183, 40–45. https://doi.org/10.1016/j.envpol.2012.11.025

Lai, P.C., Wong, C.M., Hedley, A.J., Lo, S.V., Leung, P.Y., Kong, J., Leung, G.M., 2004. Understanding the Spatial Clustering of Severe Acute Respiratory Syndrome (SARS) in Hong Kong. Environ. Health Perspect. 112, 1550–1556. https://doi.org/10.1289/ehp.7117

Leung, G.M., Hedley, A.J., Ho, L.-M., Chau, P., Wong, I.O.L., Thach, T.Q., Ghani, A.C., Donnelly, C.A., Fraser, C., Riley, S., Ferguson, N.M., Anderson, R.M., Tsang, T., Leung, P.-Y., Wong, V., Chan, J.C.K., Tsui, E., Lo, S.-V., Lam, T.-H., 2004. The Epidemiology of Severe Acute Respiratory Syndrome in the 2003 Hong Kong Epidemic: An Analysis of All 1755 Patients. Ann. Intern. Med. 141, 662. https://doi.org/10.7326/0003-4819-141-9-200411020-00006

McCarthy, P., Byrne, D., Harrisson, S., Keithley, J., 1985. Respiratory conditions: Effect of housing and other factors. J. Epidemiol. Community Health 39, 15–19. https://doi.org/10.1136/jech.39.1.15

Mogi, R., Spijker, J., 2020. The influence of social and economic ties to the spread of COVID-19 in Europe. SocArXiv.

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

Moore, D.A., Carpenter, T.E., 1999. Spatial Analytical Methods and Geographic Information Systems: Use in Health Research and Epidemiology. Epidemiol. Rev. 21, 143–161.

https://doi.org/10.1093/oxfordjournals.epirev.a017993
Niachou, K., Hassid, S., Santamouris, M., Livada, I., 2008. Experimental performance investigation of natural, mechanical and hybrid ventilation in urban environment. Build. Environ. 43, 1373–1382. https://doi.org/10.1016/j.buildenv.2007.01.046
Ogen, Y., 2020. Assessing nitrogen dioxide (NO2) levels as a contributing factor to coronavirus (COVID-19) fatality. Sci. Total Environ. 726, 138605. https://doi.org/10.1016/j.scitotenv.2020.138605
Oke, T.R., 1988. Street design and urban canopy layer climate. Energy Build. 11, 103–113. https://doi.org/10.1016/0378-7788(88)90026-6
Pareek, M., Bangash, M.N., Pareek, N., Pan, D., Sze, S., Minhas, J.S., Hanif, W., Khunti, K., 2020. Ethnicity and COVID-19: an urgent public health research priority. Lancet 395, 1421–1422. https://doi.org/10.1016/S0140-6736(20)30922-3
Penn, A., Hillier, B., Banister, D., Xu, J., 1998. Configurational modelling of urban movement networks. Environ. Plan. B Plan. Des. 25, 59–84. https://doi.org/10.1068/b250059
Qu, G., Li, X., Hu, L., Jiang, G., 2020. An Imperative Need for Research on the Role of Environmental Factors in Transmission of Novel Coronavirus (COVID-19). Environ. Sci. Technol. 54, 3730–3732. https://doi.org/10.1021/acs.est.0c01102
Rafailidis, S., 1997. Influence of building areal density and roof shape on the wind characteristics above a town. Boundary-Layer Meteorol. 85, 255–271. https://doi.org/10.1023/A:1000426316328
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. https://doi.org/10.1016/j.amepre.2020.04.003
Rocklöv, J., Sjödin, H., 2020. High population densities catalyse the spread of COVID-19. J. Travel Med. 27, taaa038.
Rothan, H.A., Byrareddy, S.N., 2020. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J. Autoimmun. 109, 102433. https://doi.org/10.1016/j.jaut.2020.102433
Ruan, Q., Yang, K., Wang, W., Jiang, L., Song, J., 2020. Clinical predictors of mortality due to

COVID-19 based on an analysis of data of 150 patients from Wuhan, China. Intensive Care Med. 46, 846–848. https://doi.org/10.1007/s00134-020-05991-x

Shen, J., Gao, S., 2008. A Solution to Separation and Multicollinearity in Multiple Logistic Regression. J. Data Sci. 6, 515–531. https://doi.org/10.6339/JDS.2008.06(4).395

Simon, N., Friedman, J., Hastie, T., Tibshirani, R., 2011. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. J. Stat. Softw. 39. https://doi.org/10.18637/jss.v039.i05

Smedje, G., Norbäck, D., 2000. New Ventilation Systems at Select Schools in Sweden—Effects on Asthma and Exposure. Arch. Environ. Health 55, 18–25. https://doi.org/10.1080/00039890009603380

Steyerberg, E.W., Eijkemans, M.J.C., Habbema, J.D.F., 1999. Stepwise selection in small data sets: A simulation study of bias in logistic regression analysis. J. Clin. Epidemiol. 52, 935–942. https://doi.org/10.1016/S0895-4356(99)00103-1

Tibshirani, R., 2011. Regression shrinkage and selection via the lasso: a retrospective. J. R. Stat. Soc. Ser. B (Statistical Methodol. 73, 273–282. https://doi.org/10.1111/J.1467-9868.2011.00771.X@10.1111/(ISSN)1467-9868.TOP_SERIES_B_RESEARCH

Tibshirani, R., 1996. Regression Shrinkage and Selection Via the Lasso. J. R. Stat. Soc. Ser. B 58, 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Vu, V.H., Le, X.Q., Pham, N.H., Hens, L., 2013. Application of GIS and modelling in health risk assessment for urban road mobility. Environ. Sci. Pollut. Res. 20, 5138–5149. https://doi.org/10.1007/s11356-013-1492-5

Wargocki, P., 2013. The Effects of Ventilation in Homes on Health. Int. J. Vent. 12, 101–118. https://doi.org/10.1080/14733315.2013.11684005

Wolf, M., 2016. Rethinking Urban Epidemiology: Natures, Networks and Materialities. Int. J. Urban Reg. Res. 40, 958–982. https://doi.org/10.1111/1468-2427.12381

Wong, M.S., Ho, H.C., Yang, L., Shi, W., Yang, J., Chan, T.C., 2017. Spatial variability of excess mortality during prolonged dust events in a high-density city: A time-stratified spatial regression approach. Int. J. Health Geogr. 16, 26. https://doi.org/10.1186/s12942-017-0099-3

Wong, M.S., Nichol, J.E., To, P.H., Wang, J., 2010. A simple method for designation of urban ventilation corridors and its application to urban heat island analysis. Build. Environ. 45, 1880–1889. https://doi.org/10.1016/j.buildenv.2010.02.019

World Health Organization, 2020a. WHO Director-General’s opening remarks at the media briefing on COVID-19 - 11 March 2020 [WWW Document]. URL https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 (accessed 6.12.20).

World Health Organization, 2020b. Coronavirus disease 2019 (COVID-19): Situation Report - 144 [WWW Document]. URL https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200612-covid-19-sitrep-144.pdf?sfvrsn=66ff9f4f_2 (accessed 6.12.20).

World Health Organization, 2003. Consensus document on the epidemiology of severe acute respiratory syndrome (SARS).

World Health Organization Regional Office for the Western Pacific, 2003. Investigation report on environmental transmission of SARS in Amoy Gardens.

Wu, X., Nethery, R.C., Sabath, B.M., Braun, D., Dominici, F., 2020. Exposure to air pollution and COVID-19 mortality in the United States. medRxiv.

Yang, J., Jin, S., Xiao, X., Jin, C., Xia, J. (Cecilia), Li, X., Wang, S., 2019. Local climate zone ventilation and urban land surface temperatures: Towards a performance-based and wind-sensitive planning proposal in megacities. Sustain. Cities Soc. 47, 101487. https://doi.org/10.1016/j.scs.2019.101487

Yang, J., Wong, M.S., Menenti, M., Nichol, J., 2015. Modeling the effective emissivity of the urban canopy using sky view factor. ISPRS J. Photogramm. Remote Sens. 105, 211–219. https://doi.org/10.1016/j.isprsjprs.2015.04.006

Yang, J., Wong, M.S., Menenti, M., Nichol, J., Voogt, J., Krayenhoff, E.S., Chan, P.W., 2016. Development of an improved urban emissivity model based on sky view factor for retrieving effective emissivity and surface temperature over urban areas. ISPRS J. Photogramm. Remote Sens. 122, 30–40. https://doi.org/10.1016/j.isprsjprs.2016.09.007

Yang, X., Li, Y., 2015. The impact of building density and building height heterogeneity on average urban albedo and street surface temperature. Build. Environ. 90, 146–156.

https://doi.org/10.1016/j.buildenv.2015.03.037

Yu, I.T.S., Li, Y., Wong, T.W., Tam, W., Chan, A.T., Lee, J.H.W., Leung, D.Y.C., Ho, T., 2004. Evidence of Airborne Transmission of the Severe Acute Respiratory Syndrome Virus. N. Engl. J. Med. 350, 1731–1739. https://doi.org/10.1056/NEJMoa032867

Zhang, P., Atkinson, P.M., 2008. Modelling the effect of urbanization on the transmission of an infectious disease. Math. Biosci. 211, 166–185. https://doi.org/10.1016/j.mbs.2007.10.007

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G.F., Tan, W., 2020. A novel coronavirus from patients with pneumonia in China, 2019. N. Engl. J. Med. 382, 727–733. https://doi.org/10.1056/NEJMoa2001017

Zhu, Y., Xie, J., Huang, F., Cao, L., 2020. Association between short-term exposure to air pollution and COVID-19 infection: Evidence from China. Sci. Total Environ. 727, 138704. https://doi.org/10.1016/j.scitotenv.2020.138704