Forecasting COVID-19 onset risk and evaluating spatiotemporal variations of the lockdown effect in China

Wenzhong Shi1*, Chengzhuo Tong1, Anshu Zhang1, Bin Wang2, Zhicheng Shi1, Yepeng Yao1, Peng Jia3,4

1Laboratory for Smart City and Spatial Big Data Analytics, Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China
2College of Oceanography and Space Informatics, China University of Petroleum, Qingdao, China
3International Initiative on Spatial Lifecourse Epidemiology (ISLE), Hong Kong, China
4Faculty of Geo-information Science and Earth Observation, University of Twente, Enschede, The Netherlands

*Correspondence author. Email: lswzshi@polyu.edu.hk

Abstract: It is important to forecast the risk of COVID-19 symptom onset and thereby evaluate how effectively the city lockdown measure could reduce this risk. This study is a first comprehensive, high-resolution investigation of spatiotemporal heterogeneities in the effect of the Wuhan lockdown on the risk of COVID-19 symptom onset in all 347 Chinese cities. An extended Weight Kernel Density Estimation model was developed to predict the COVID-19 onset risk under two scenarios (i.e., with and without Wuhan lockdown). The Wuhan lockdown, compared with the scenario without lockdown implementation, delayed the arrival of the COVID-19 onset risk peak for 1-2 days in general and lowered risk peak values among all cities. The decrease of the onset risk attributed to the lockdown was more than 8% in over 40% of
Chinese cities, and up to 21.3% in some cities. Lockdown was the most effective in areas with medium risk before lockdown.

**Introduction**

Emerging infectious diseases, such as coronavirus disease 2019 (COVID-19), have become global challenges for the public health sector\(^1\,^2\). This new virus is highly contagious and can be transmitted through respiratory droplets or physical contact\(^3\,^4\). As of 8th April, 2020, 1,353,361 persons worldwide have been identified as being infected by COVID-19\(^5\). Confirmed cases have appeared in 213 countries or regions\(^5\), and currently, the number is still rapidly increasing (Fig. 1). With the fast development of modern transportation, people are able to travel on an accelerating scale and speed, which could make pathogens spread more easily and hence, greatly aggravate the transmission of diseases. To reduce the spread of COVID-19, China, from 23rd January 2020, imposed a lockdown measure in Wuhan (the regional epicentre of the COVID-19 pandemic at that time), the capital of Hubei Province with a population over 10 million. The Wuhan lockdown prevented anyone from entering or leaving Wuhan by any means of transportation from that specific lockdown moment. It also limited local movements of residents to only those critical activities. As a result, from around 10th February 2020, just 18 days after the implementation of the lockdown, a significant drop in daily new COVID-19 infections was witnessed in all of China’s provinces except Hubei. Such early decreases in new infections enabled the relatively limited healthcare resources to be saved for both COVID-19 and those non-COVID-19 patients in the greatest need across China, which has further enabled the resource reallocation to COVID-19 patients in Wuhan and speeded up the recovery of the whole
country from the epidemic. As of 8th April 2020, the end of the Wuhan lockdown, 77,838 (about 93%) confirmed cases across China have recovered.

![Spatial distribution of COVID-19 confirmed cases worldwide by 8th April 2020.](image)

**Fig. 1.** Spatial distribution of COVID-19 confirmed cases worldwide by 8th April 2020. The names of four countries with the highest numbers of confirmed cases are shown. The numbers of confirmed cases were collected from the World Health Organization’s COVID-19 Situation Report-79.

Lockdown increasingly seems a necessary measure to curb the current escalating pandemic for many other countries, thus it is urgent to provide scientific evidence regarding the exact effect of the lockdown measure and precisely how effectively the lockdown measure could bring COVID-19 under control. Some early efforts have been made in this direction, however the findings are mixed. For example, Yang et al. (2020) showed that, if Wuhan had been locked
down five days later, the cumulative number of COVID-19 infections during 23rd January-24th April 2020 would have tripled in China. Wu et al. (2020) predicted that, even with the lockdown implemented in Wuhan, many other major cities in China would still undergo local outbreaks with exponentially growing infections, as seen in Wuhan.

A recent study estimated that the Wuhan lockdown was associated with the later appearance of the reported COVID-19 cases in other cities by 2.91 days (95% CI: 2.54-3.29). Nevertheless, all existing studies have used mathematical models of infectious disease transmission that rely on theoretical epidemiological parameters for making prediction (e.g., Susceptible-Exposed-Infectious-Recovered [SEIR] models). In such instances, the spatial relationships among cities (i.e., the First Law of Geography that everything is related to everything else, but near things are more related than distant things) have been significantly downplayed. Facing new infectious diseases (e.g., COVID-19) with limited prior knowledge of their features and also limited associated comparability with previous diseases, there has been much uncertainty in setting theoretical parameters and assumptions of mathematical prediction models, which could have led to highly mixed and uncertain conclusions among existing studies. Moreover, all existing models, by using daily confirmed COVID-19 cases, are based on and also predictive of only the date of reporting, which is usually later than the date of the COVID-19 symptom onset. It has been found that COVID-19 patients are most infectious during the first week after the symptom onset. Therefore, by using findings purely on the basis of COVID-19 dates, the best time for preventing new COVID-19 infections may be missed. Infectious diseases are spread by the spatial movements of dynamic hosts and/or vectors, thus adopting appropriate spatial, data-driven models under the Fourth Paradigm has been considered a minimum requirement for exploring the progression of the risk of new infectious diseases such
as COVID-19, and also, for evaluating the effects of epidemic containment measures (e.g., Wuhan lockdown)\(^{19}\). The Weight Kernel Density Estimation (WKDE) model is one of such models, which conducts retrospective analyses to speculate the dates of infection for confirmed cases, on the basis of their locational information, and predicts the place-specific risk of infection caused by spatial movements of infected people\(^{20}\). Such spatial, data-driven models have integrated the First Law of Geography and alleviated reliance on theoretical assumptions and parameters, thereby making more robust, place-specific predictions than those based on mathematical prediction models in the contexts of new infectious diseases. The WKDE model, however, did not consider the changing effects of modern factors on traditional spatial relationships among places, such as travel vehicles, which have decreased travel impedance among places. For more accurate prediction, it is necessary to use dynamic mobility data to capture those modern factors, and improve existing spatial models to incorporate such big data into prediction.

In this study, an extended WKDE model has been developed, on the basis of the original WKDE model, to forecast the risk of COVID-19 symptom onset and evaluate the effects of the Wuhan lockdown on the risk of the COVID-19 onset at a high spatiotemporal resolution across China. This is a first-ever comprehensive, high-resolution investigation of spatiotemporal heterogeneities in the effects of the Wuhan lockdown on the risk of the COVID-19 onset in all 347 Chinese cities. The extended WKDE incorporates inter-city human mobility data for the calibration of traditional spatial relationships among cities in a high-speed era, to predict the risk of the COVID-19 onset and the spatiotemporal patterns of the onset risk under two scenarios — with and without Wuhan lockdown. The analysis was based on a spatiotemporal dataset of 40,486 confirmed COVID-19 cases in China during the period from the 31st December 2019 to
2nd March 2020 (the 40th day after Wuhan lockdown), among which 1,189 cases had available reported dates of symptom onset; dates of symptom onset for the remaining 39,297 cases without such information were speculated by an established statistical method. The daily Wuhan out-migration to all other cities and the associated percentage of Wuhan migrants to every other city were calculated to indicate inter-city human movements in the extended WKDE model.

Results

Spatiotemporal evolution of the risk of COVID-19 symptom onset. The risk of the COVID-19 symptom onset during the study period under the city lockdown scenario which happened in reality was first estimated by the extended WKDE model, with the spatiotemporal variations of the risk demonstrated. On 25th December 2019, areas with the risk of COVID-19 symptom onset higher than 0.8 were only in Wuhan and surrounding cities, which were largely in eastern Hubei Province; the onset risk in 96.8% of the areas in China was lower than 0.2 (Fig. 2a). The influences of human mobility on the onset risk were not apparent until approximately one week before the Wuhan lockdown, in which cities receiving a large number of passengers from Wuhan started to present higher onset risk, such as Chongqing (west), Beijing (north), and Guangzhou (south) (Fig. 2c). Approximately one week after the Wuhan lockdown, the onset risk reached the peak, before decreasing steadily over the whole area outside Hubei Province (Fig. 2f). The risks of symptom onset after two (Fig. 2g) and three weeks of Wuhan lockdown (Fig. 2h), in about 64.3% and 81.6% of cities outside Hubei Province, dropped to below 0.6, respectively. By the end of the fifth week following the Wuhan lockdown, the areas with the risk of onset higher than 0.8 shrank to only Wuhan and surrounding cities in Hubei (Fig. 2k).
Dec 25, 2019

Jan 05, 2020

Jan 15, 2020

Jan 20, 2020

Jan 25, 2020 (lockdown)

Jan 30, 2020 (lockdown)

Feb 05, 2020 (lockdown)

Feb 10, 2020

Feb 15, 2020

Feb 27, 2020

Jan 25, 2020 (non-lockdown)

Jan 30, 2020 (non-lockdown)

Feb 05, 2020 (non-lockdown)
Fig. 2. Predicted risk of COVID-19 symptom onset across all Chinese cities before the date of Wuhan lockdown (a-d) and under two scenarios – with (e-j) and without (k-m) Wuhan lockdown measure – after the date of Wuhan lockdown. The predicted risk of COVID-19 symptom onset were resulted from the extended Weight Kernel Density Estimation (WKDE) model, on the basis of data for the confirmed cases with locations where they had a period of study prior to the diagnosis and inter-city human mobility data. The predictions under two scenarios were made by differing the human mobility intensity from Wuhan to other cities after the date of Wuhan lockdown: the human mobility intensity was regarded as zero under the lockdown scenario, and was estimated by the human mobility intensity on the corresponding time period in 2019 under the non-lockdown scenario (see Methods for details).

The accuracy of the predicted risk of symptom onset was evaluated on a daily basis, by calculating the percentage of confirmed cases reported in areas in which the predicted risk of symptom onset was higher than 0.8. A percentage higher than 70% is considered acceptable. The extended WKDE model resulted in prediction accuracy of over 70% regarding the onset risk in the first week after the base day. Such accuracy achieved by the extended WKDE model was higher than that resulting from the original WKDE model (Fig. 3). The prediction accuracy in the second week after the base day, as possibly could be expected, was naturally lower due to the accumulation of prediction errors over time. The outperformance of the extended WKDE model was attributed to human mobility being integrated into the WKDE model.
Fig. 3. Accuracy of the predicted risk of COVID-19 symptom onset by the extended and original WKDE models. The time interval denotes the period between the base date and the date of prediction (i.e., the date on which the risk of symptom onset is predicted). The horizontal line in the box denotes the median, and the lower and upper edges of the box represent the second and third quartile, respectively; the lines emanating from the box upwards and downwards represent the maximum and minimum values, respectively. The predicted onset risk was a standardized value between 0 and 1, indicating the risk relative to the highest predicted risk at all locations on the date of prediction. The accuracy of prediction is measured by the percentage of actual confirmed cases on the date of prediction that occur in the areas with predicted onset risk higher than 0.820.

Spatiotemporal patterns of effects of Wuhan lockdown on COVID-19 onset risk. The effects of the Wuhan lockdown on the risk of the COVID-19 symptom onset were then evaluated by predicting the risk of the COVID-19 onset under two scenarios — with and without Wuhan lockdown, with the used of the extended WKDE model. Compared to the COVID-19 onset risk under the lockdown scenario (Fig. 2e-g), the onset risk under the non-lockdown scenario on the same date was significantly higher (Fig. 2k-m). Around the particular time that the onset risk reached a peak (Fig. 2m), the areas of onset risk which was higher than 0.8 would have expanded
to include all Chinese cities, except those in the north and west, which had a low inter-city population flows from Wuhan. Wuhan lockdown has, indeed, contributed to the decreased risk of the COVID-19 onset in all 347 cities of China (Fig. 4a). The decrease in the onset risk attributed to Wuhan lockdown was up to 21.3% throughout all cities, with more than 8%, 12%, and 16% in 146, 58, and 28 cities, respectively. Note that these decreases were the most conservative estimates, as the predicted risk of onset under the non-lockdown scenario was made based on the confirmed cases under the lockdown scenario. Most if not all cities, without the Wuhan lockdown, in theory, would have been likely to have had more than the currently recorded cases. Therefore, compared to the risk of onset without Wuhan lockdown, the actual effect of this lockdown on the spread of COVID-19 (i.e., subsequent risk decrease of COVID-19 symptom onset) should undoubtedly have been greater.

A daily comparison of the predicted onset risk under lockdown and non-lockdown scenarios reflected the contributions by the Wuhan lockdown in three aspects: a constant lower daily onset risk after the lockdown, the delayed arrival of peaks of the daily onset risk by 1-2 days, and the subsequent lower peak risks. These three contributions could be observed not only in megacities, such as Shanghai, Beijing, and Shenzhen, but also in medium-size cities, such as Xiangtan (in Hunan Province), Hanzhong (Shaanxi), Zhangzhou (Fujian) (Fig. 4b). After the Wuhan lockdown had subsequently prevented more disease cases to be imported to other cities, the majority of cities outside Hubei were able to control the local spread of the disease, given the drops of the onset risk to an apparently lower level after two weeks (Fig. 4b). About three weeks after the Wuhan lockdown, decreases in the risk of onset relative to the corresponding peak values ranged from 21.1% to 78.9% among 285 cities outside Hubei with a peak risk of above
0.4, from 15.7% to 62.4% among 27 cities with a peak risk between 0.2-0.4, and from 0.2% to 5.1% in 18 cities with a peak risk of below 0.2 (Fig. 4a).

Fig. 4. The risk of COVID-19 symptom onset under two scenarios – with and without Wuhan lockdown measure – from 24th January to 5th February 2020. (a) Average daily percentage reduction in the onset risk in all Chinese cities in the lockdown scenario, compared with the non-lockdown scenario. (b) The onset risk under two scenarios in six selected cities (three megacities and three medium-size cities). The plotted values were computed from the predicted risk of COVID-19 symptom onset under the two scenarios resultant from the extended WKDE model, as described in the legend of Fig. 3.
The reduction of the onset risks in different cities attributed to the Wuhan lockdown were heterogeneous and related to two factors: the intensity of human mobility from Wuhan to individual other cities, and the level of existing risk of onset in individual cities by the date of the Wuhan lockdown. Most of western and northeastern China showed a low reduction percentage (<4%) of the onset of risk attributed to the Wuhan lockdown (Fig. 4a). The human mobility from Wuhan to those areas was much less intense than to most other cities in China (Fig. 5). In the six selected cities which were deemed representative of all Chinese cities in terms of the geography and economic development, it was also observed that the higher the existing risk of onset by the lockdown date, the less the percentage of the onset risk reduction was attributed to the Wuhan lockdown (Fig. 4b). In the whole country, areas with a high percentage (>8%) of reduction of onset risk attributed to the Wuhan lockdown (Fig. 4b) largely coincided with areas with medium risk of onset (0.2-0.6) before the lockdown (Fig. 2d).

**Discussion**

This study for the first time has provided data-driven evidence on the effects of the Wuhan lockdown measures on the risk of COVID-19 symptom onset in all 347 Chinese cities in a high spatiotemporal resolution, i.e., demonstrating the changing patterns of the risk of COVID-19 symptom onset at city level on a daily basis. Specifically, Wuhan lockdown has lowered and delayed (for 1-2 days) the arrival of the daily onset risk peak in all other cities, also resulting in decreased risk of symptom onset in those cities following the Wuhan lockdown on 23rd January 2020 when compared to the scenario without the lockdown. It has also reduced the infection risk by imported cases from Wuhan in other cities within one week after the lockdown was imposed. This situation varied across cities: the larger the volume of passengers a city received from
Wuhan, the larger the avoided risk of symptom onset. Furthermore, the Wuhan lockdown was found to be most effective in reducing the onset risk in areas with medium risks prior to lockdown.

Findings from this study have important and unique implications for public health and especially epidemic response. They would serve as strong place-specific evidence regarding the effectiveness and efficiency of the measure of city lockdown in reducing the COVID-19 onset risk. Results presented in this study can also be supported by those from a recent study, which concluded from the results of a SEIR model that the Wuhan lockdown delayed the arrival of COVID-19 in other cities by 2.91 days\textsuperscript{10}. However, the delayed arrival presented in that study, in reality, referred to the ‘delayed reporting’, as only the reported dates of confirmed cases were used. This current study has predicted the risk of symptom onset, which is conceptually less affected by some uncertain factors (e.g., time between presentation of symptoms and tests, diagnosis, and reporting). Such predictions could inform public health agencies ahead of catastrophic spread and hence guide local epidemic control and subsequently precise prevention efforts. Beyond early warnings, this study sheds further light on spatiotemporal heterogeneities in the predicted risks, which would enable precision disease control and prevention by pointing out not only what to do and when to do, but also where and how to put these findings into practice, such as the identification of the exact areas of high risk and the level of risk, thus to better enable decisions concerning precise interventions.

This study does have some limitations. Firstly, only 40,486 confirmed cases with community-level locations were used rather than all 75,465 confirmed cases revealed during the study period in China. The incompleteness of such exposure history collected from confirmed cases could affect the results. Such effects, however, are considered limited toward our study
aim, which focuses more on the areas outside Hubei Province. Cases without detailed locations mainly took place in severely affected cities, such as Wuhan, in which people are intensively engaged in treatments and quarantine, hence, with limited capacity of committing to such epidemiological data collection efforts. Secondly, the method of speculating dates of symptom onset used in this study, although officially documented, may still be subject to uncertainties, which needs to be further improved in the future. Thirdly, COVID-19 infectiousness and susceptibility in different populations were not considered in the current model, due to incomplete demographical information. Although this is considered acceptable as the travel and contact history has been playing a major role in the COVID-19 infection, more detailed demographical information and confirmed findings on varying infectiousness and susceptibility from clinical studies would further improve the quality of the model. Lastly, the current model considers only out-migrants from Wuhan without including travel flows between other cities, although the latter has been largely minimized within the study period due to a high tendency to avoid COVID-19 and lockdown measures implemented to a different extent across China.

Nevertheless, despite the limitations listed above, this study has presented an unprecedented spatiotemporal prediction, thanks to the detailed publicly available information gathered from anonymous cases. Transparent reporting of travel and contact history of such a large number of anonymous infected cases has been realized in China, thus opening a new avenue in the era of big data under the Fourth Paradigm, for more advanced models to refine results from mathematical prediction models. It also enables and encourages multiple stakeholders outside the public health sector to be involved in collaborative control and prevention efforts to contribute to the curbing of increasingly more frequent and complex epidemics in the future\textsuperscript{23,24}. 
The high spatiotemporal resolution evidence from this study would be necessary for lockdown decision-making not only in all countries undergoing the first wave of infections, but in African and Latin American countries that will inevitably be hit by the next wave of infections\textsuperscript{25}. More importantly, the extended WKDE model can be automatically fed by real-time spatiotemporal surveillance data from multiple sources to make robust, timely predictions, which is difficult to be realized in traditional approaches. Such predictions would be a core component of epidemic early warning systems regarding the prevention of the future epidemics.

**Methods**

**Data sources.** Spatiotemporal data of 40,486 confirmed COVID-19 cases in China during the period from the 31\textsuperscript{st} December 2019 to 2\textsuperscript{nd} March 2020 (the 40\textsuperscript{th} day after Wuhan lockdown), among a total of 75,465 confirmed cases during this period\textsuperscript{21}, were collected from official reports by provincial and municipal health commissions, and from supplemental information in the public media, such as People's Daily, Tencent Health, and Baidu Map. These cases had available dates of reporting and self-reported, community-level locations where they had a period of stay prior to diagnosis. Among them, 1,189 cases had available reported dates of symptom onset, with the earliest onset of symptom dated on 8\textsuperscript{th} December 2019\textsuperscript{21}.

Two human mobility parameters derived from massive positioning service data in China on Baidu Map platform\textsuperscript{26} were obtained during January-March 2019 and 2020: the daily Baidu out-migration index for Wuhan, representing the magnitude of the total population leaving Wuhan to other cities daily, and the associated percentage of migrants from Wuhan to every other city (Fig. 5). Also obtained were the daily passenger loads of all flights, railways and bus services from Wuhan to every other city during December 2019 to March 2020\textsuperscript{27-29}. 

15
Fig. 5. The daily percentage of migrants from Wuhan to every other city during the study period. The figure was based on the massive positioning service data on the Baidu Map platform available on https://qianxi.baidu.com/26. Several hotspot destinations of passengers departing from Wuhan, as reflected from this figure, are labelled.

**Daily migration.** Two indicators of daily migration during the study period were calculated: the daily Wuhan out-migration to all other cities and the daily percentage of Wuhan migrants to every other city. Both indicators since 1st January 2020 could be directly obtained from Baidu migration data, which, however, were not available for December 2019. Therefore, both indicators for December 2019 were calculated based on Baidu data in early January 2020 and the daily passenger load data, under the assumption that the daily passenger load factor was stable during December 2019 and early January 2020. The daily Wuhan out-migration for December
2019 was calculated by averaging the ratios of the daily Baidu out-migration index to the daily passenger loads in early January 2020, followed by multiplying the average ratio by the daily passenger loads leaving Wuhan to other cities in December 2019. The daily percentage of migrants from Wuhan to other cities in December 2019 was estimated as the ratio of the daily passenger load from Wuhan to that city to the daily passenger loads from Wuhan to all other cities. This method was also used to calculate the daily percentage of migrants in 2020 from Wuhan to some cities receiving a considerably small number of migrants from Wuhan, where Baidu migration data were suppressed from reporting for confidentiality reasons.

**Date of symptom onset for COVID-19 cases.** An established statistical method was adopted to speculate dates of symptom onset for 39,297 confirmed cases without such information, having been described elsewhere. The method results in a probability $p_{\text{waiting}}(\Delta t)$ that each confirmed case has a waiting time of $\Delta t$ day(s) between onset and diagnosis. Among $m$ confirmed cases in the same city on the same day, a total of $m \cdot p_{\text{waiting}}(\Delta t)$ confirmed cases were randomly selected and assigned a waiting time of $\Delta t$. As a result, the distribution of dates of symptom onset among the 40,486 confirmed cases used in this study was consistent with that of actual dates of symptom onset among all 75,465 confirmed cases reported by the World Health Organization (Fig. 6).
Fig. 6. Distribution of the speculated dates of COVID-19 symptom onset among 40,486 confirmed cases used in this study, and of the actual dates of COVID-19 symptom onset among 75,465 confirmed cases reported by the World Health Organization as of 20th February 2020. Data on 40,486 confirmed cases in China from 31st December 2019 to 2nd March 2020 were collected from official reports by provincial and municipal health commissions and public media. These cases had available dates of reporting and community-level locations where they had a period of stay prior to diagnosis. Among them, 1,189 cases had available reported dates of symptom onset. An established statistical method\textsuperscript{22} was used to speculate the dates when the onset dates were unavailable. The 75,465 confirmed cases were reported by the WHO\textsuperscript{21} and did not included the fine-scale location of the cases needed in this study.

**Extended WKDE model for predicting the risk of COVID-19 onset.** The model makes the prediction through the following three steps:

**Step 1:** Retrospective inference on historical existence likelihood of COVID-19 infection at each location, in which each confirmed case had a period of stay. The aim of this step is to speculate the infection date of each confirmed case, on the basis of the date of symptom onset, for estimating the risk that the case has transmitted pathogens to others from the infection date.
The incubation period of each confirmed case from infection to symptom onset is modelled following a Weibull distribution\textsuperscript{30}:

\[
    p_{\text{incubation}}(\Delta t) = k \lambda^{-k} \Delta t^{k-1} e^{-((\Delta t/\lambda)^{k})}
\]

(1)

where \(p_{\text{incubation}}(\Delta t)\) denotes the probability that the incubation period of each confirmed case equals to \(\Delta t\) days; \(\lambda\) and \(k\) denote the mean and standard deviation of the incubation period, which, in this study, were assumed to be 5.2 and 2.8 days, respectively\textsuperscript{31,32}; and \(e\) denotes the natural exponential.

All days in this study period (8\textsuperscript{th} December 2019 - 27\textsuperscript{th} February 2020) are in the order denoted as \(t_1, t_2, \ldots \ t_{82}\). The probability that each confirmed case was infected on a certain day and thus became infectious is:

\[
    p_{\text{infection}}(L, t_i) = 1 - \prod_{t_L > t_i} (1 - p_{\text{incubation}}(t_L - t_i))^{n(t_L)}
\]

(2)

where \(p_{\text{infection}}(L, t_i)\) denotes the probability that one confirmed case at location \(L\) was infected on day \(t_i\); \(t_L\) denotes the day of symptom onset for the confirmed case at location \(L\); \(n(t_L)\) is the number of onset cases at location \(L\) on day \(t_L\); \(p_{\text{incubation}}(t_L - t_i)\) denotes the probability that the incubation period of the confirmed case is equal to \((t_L - t_i)\) days.

**Step 2:** Spatial extrapolation for inferring historical existence likelihood of COVID-19 infection in the whole study area. Let \(L_1, L_2, \ldots, L_n\) be the unique locations of all confirmed cases used for the prediction. The risk of infection at each random location is estimated as:

\[
    p_{\text{infection}}(S, t_i) = n(t_i)^{-1} \sum_{j=1}^{n(t_i)} M(S, t_i) p_{\text{infection}}(L_j, t_i) K_h(S - L_j)
\]

(3)

where \(p_{\text{infection}}(S, t_i)\) denotes the probability that any infected person visited a random location \(S\), on day \(t_i\) and posed the risk of infection to others nearby; \(L_j\) denotes the \(j\)-th location among unique locations of all the confirmed cases; \(K_h(S - L_j)\) denotes a Gaussian kernel between locations \(S\) and \(L_j\).
\[ K_h(S - L_j) = \frac{1}{2\pi \sqrt{\det(h)}} \exp\left( -\frac{1}{2} (S - L_j)^T h^{-1}(S - L_j) \right) \] (4)

where \( h \) denotes the bandwidth matrix equating to \( n^{-1/2} \hat{\Sigma} \), with \( n \) being the total number of confirmed cases in all 82 days, and \( \hat{\Sigma} \) being the covariance matrix of \((x_1, \ldots, x_n)\) and \((y_1, \ldots, y_n)\), denoting the vectors of \( x \) and \( y \) coordinates of \( L_1, L_2, \ldots, L_n \); \( M(S, t) \) denotes a human mobility factor at location \( S \), on day \( t_i \), calculated as:

\[
M(S, t_i) = \begin{cases} 
  i^{-1} \sum_{k=1}^i b_{sk} V_k, & \text{S is outside Wuhan} \\
  1, & \text{S is inside Wuhan}
\end{cases} \] (5)

where \( b_{sk} \) denotes the daily percentage of Wuhan migrants to the city containing location \( S \), on day \( t_k \); \( V_k \) denotes the daily Wuhan out-migration to other cities on day \( t_k \) prior to \( t_i \). Note that \( M(S, t) \) after the lockdown day \( t_{47} \) is calculated differently under two scenarios, i.e., with and without Wuhan lockdown measures implemented. \( M(S, t) \) after \( t_{47} \) is equal to \( M(S, t_{47}) \) under the lockdown scenario because it was assumed no one left Wuhan after lockdown. Under the non-lockdown scenario, \( M(S, t) \) after \( t_{47} \) is computed by assigning \( b_S \) and \( V_k \) the corresponding values on the same lunar calendar date in 2019 as \( t_k \), because of the similar migration patterns during that period over two years\(^{10}\).

**Step 3:** Making predictions of the risk of COVID-19 onset at each random location on a specific day in near future:

\[
P_{\text{onset}}(S, t_z) = 1 - \prod_{t_i \leq t_z} (1 - P_{\text{infection}}(S, t_i) P_{\text{incubation}}(t_z - t_i)) \] (6)

where \( P_{\text{onset}}(S, t_z) \) denotes the likelihood that at least one person infected by a confirmed case at location \( S \) develops clinical symptoms on day \( t_z \); \( t_i \) denotes the date of infection for that person, so always \( i < z \). Note that \( P_{\text{onset}}(S, t_z) \) values represent point estimates over the continuous space. Such a risk measure may not be intuitive for decision making and hence could be standardized to
the range of 0 and 1 by being divided by the maximum predicted risk among all locations on day $t_2$. By doing so, the standardized predicted risk is seen as the relative risk of symptom onset to the highest risk of symptom onset in the study area, which can serve as an intuitive indicator for epidemic control and prevention work on the ground. Moreover, point estimates of the risk can be averaged flexibly over any areal unit (e.g., city, residential community), and hence could overcome the modifiable areal unit problem (MAUP) during epidemic response.\textsuperscript{33}

Assessment of prediction models. The accuracy of the predicted risk of symptom onset has been evaluated on a daily basis, by calculating the percentage of confirmed cases reported in areas in which the predicted risk of symptom onset was higher than 0.8. A percentage higher than 70\% is considered acceptable.\textsuperscript{20} The predicted results were also compared to results from the original WKDE model\textsuperscript{20} for all cities. Since prediction errors could accumulate with time in a prediction for further future, the risk predicted on the basis of data no later than “the day before” is usually the most accurate. As a consequence, all risks of symptom onset mentioned in this study have been predicted based on confirmed COVID-19 cases with onset dates no later than the previous day, except the risk on 27\textsuperscript{th} February 2020, which was predicted based on the data on or before 20\textsuperscript{th} February 2020 (the last date of symptom onset speculated among confirmed cases).

Data availability

The datasets generated and analysed during this study are available from the corresponding author on reasonable request.
Code availability

The MATLAB scripts developed and used in this study are available from the corresponding author on reasonable request.

References

1. Shoemaker, T. R. et al. Impact of enhanced viral haemorrhagic fever surveillance on outbreak detection and response in Uganda. *Lancet Infect. Dis.* 18, 373–375 (2018).
2. Paixão, E. S., Teixeira, M. G. & Rodrigues, L. C. Zika, chikungunya and dengue: The causes and threats of new and reemerging arboviral diseases. *BMJ Glob. Heal.* 2, e000530 (2018).
3. Xia, W. et al. Clinical and CT features in pediatric patients with COVID-19 infection: Different points from adults. *Pediatr. Pulmonol.* 55, 1169–1174 (2020).
4. Wang, L., Wang, Y., Ye, D. & Liu, Q. A review of the 2019 Novel Coronavirus (COVID-19) based on current evidence. *Int. J. Antimicrob. Agents* 105948, in press (2020).
5. WHO. *Coronavirus disease 2019 (COVID-19) Situation Report-79.* https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200408-sitrep-79-covid-19.pdf?sfvrsn=4796b143_6 (2020).
6. National Health Commission of the People’s Republic of China. *Coronavirus disease 2019 (COVID-19) Situation Report-08/04/2020.* http://www.nhc.gov.cn/xcs/yqtb/202004/fa7bb40a7fbf4b2c8f3989d512fe5b77.shtml (2020).
7. Cyranoski, D. WHAT CHINA’S CORONAVIRUS RESPONSE CAN TEACH THE REST OF THE WORLD. *Nature* 579, 479–480 (2020).
8. Yang, Z. et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J. Thorac. Dis.* 12, 165-174 (2020).
9. Wu, J. T., Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* 395, 689-697 (2020).
10. Tian, H. et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science Science* 6105, eabb6105 (2020).
11. Tobler, W. R. A Computer Movie Simulating Urban Growth in the Detroit Region. *Econ. Geogr.* 46, 234–240 (1970).
12. Zhao, Y. et al. A new Seasonal Difference Space-Time Autoregressive Integrated Moving Average (SD-STARIMA) model and spatiotemporal trend prediction analysis for Hemorrhagic Fever with Renal Syndrome (HFRS). *PLoS One* 13, 1–20 (2018).
13. Myer, M. H. & Johnston, J. M. Spatiotemporal Bayesian modeling of West Nile virus: Identifying risk of infection in mosquitoes with local-scale predictors. *Sci. Total Environ.* 650, 2818–2829 (2019).
14. Liu, S. et al. Predicting the outbreak of hand, foot, and mouth disease in Nanjing, China: a time-series model based on weather variability. *Int. J. Biometeorol.* 62, 565–574 (2018).
15. Ak, Ç., Ergönül, Ö., Şencan, İ., Torunoğlu, M. A. & Gönen, M. Spatiotemporal prediction of infectious diseases using structured Gaussian processes with application to Crimean–Congo hemorrhagic fever. *PLoS Negl. Trop. Dis.* **12**, 1–20 (2018).

16. To, K. K. *et al.* Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2 : an observational cohort study. *Lancet Infect. Dis.* **3099**, 1–10 (2020).

17. Wölfel, R. *et al.* Virological assessment of hospitalized patients with COVID-2019. *Nature*, published online April 1 (2020).

18. Jia, P. *et al.* Spatial lifecourse epidemiology and infectious disease research. *Trends Parasitol.* **36**, 235–238 (2020).

19. Jia, P. *et al.* Time to spatialize epidemiology in China. *Lancet Glob. Heal.* **8**, in press (2020).

20. Bin, W. Analyzing and predicting risks of infectious diseases by geographic information science. (The Hong Kong Polytechnic University, 2015).

21. WHO. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19) (2020).

22. Shi, Y. *et al.* Using a discrete stochastic model to study the epidemic dynamics of COVID19 in Hubei, China. *J. Univ. Chinese Acad. Sci.* **37**, 145–154 (2020).

23. Bedford, J. *et al.* A new twenty-first century science for effective epidemic response. *Nature* **575**, 130–136 (2019).

24. Jia, P. *et al.* “Are we ready for a new era of high-impact and high-frequency epidemics?”. *Nature* **580**, 580(7803):321(2020).

25. Lancet. COVID-19 : learning from experience. *Lancet* **395**, 1011 (2020).

26. Baidu. Baidu migration index. https://qianxi.baidu.com/ (2020).

27. Ltd. T. co. G. Daily Flights from Wuhan. https://www.ctrip.com/ (2020).

28. CHINA ACADEMY OF RAILWAY SCIENCES CORPORATION LIMITED. 12306 China Railway. https://www.12306.cn/index/ (2020).

29. Communication & Transportation Association of China. Daily Bus from Wuhan. https://m.changtu.com/city/wuhanshi/ (2020).

30. J A Backer, D Klinkenberg, J. W. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020. *Eurosurveillance* **25**, pii=2000062 (2020).

31. Ganyani, T. *et al.* Estimating the generation interval for COVID-19 based on symptom onset data. *medRxiv* 2020.03.05.20031815 (2020).

32. Zhang, J. *et al.* Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province , China : a descriptive and modelling study. *Lancet Infect. Dis.* **3099**, 1–10 (2020).

33. S.Openshaw. *The Modifiable Areal Unit Problem*. (Geo Books, 1984).
Acknowledgements

This study was supported by The Hong Kong Polytechnic University (1-ZVN6, 1-99XK).

Author contributions

W.S. conceived and designed the study, interpreted the results, and helped develop the computation models, analyse the data, and write the manuscript. C.T. collected the data, developed the computation models, wrote the manuscript, and helped interpreted the results. A.Z. wrote the manuscript, and helped analyse the data and interpret the results. B.W. developed the computation models. Z.S. helped collect and analyse the data. Y.Y. helped analyse the data. P.J. interpreted the results, helped analyse the data, and helped write the manuscript.

Competing interests

The authors declare no competing interests.