Geography and Sustainability 1 (2020) 77–87

Contents lists available at ScienceDirect

Geography and Sustainability

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

Article

COVID-19: Challenges to GIS with Big Data

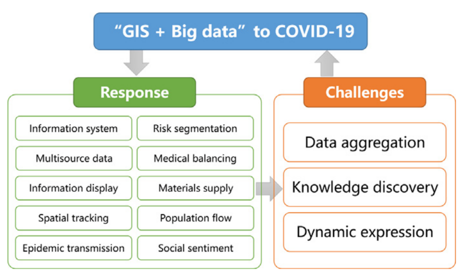
Chenghu Zhou [a,c,†,*], Fenzhen Su [a,c,e,†,*], Tao Pei [a,c], An Zhang [a], Yunyan Du [a,c], Bin Luo [a], Zhidong Cao [b,d], Juanle Wang [a,c], Wen Yuan [a], Yunqiang Zhu [a,c], Ci Song [a], Jie Chen [a], Jun Xu [a,e], Fujia Li [a], Ting Ma [a,c], Lili Jiang [a], Fengqin Yan [a], Jiawei Yi [a], Yunfeng Hu [a], Yilan Liao [a], Han Xiao [a,c]

[a] State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China
[b] Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[c] College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China
[d] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China
[e] Commission on Geographical Information Science, International Geographic Union, Beijing 100101, China

HIGHLIGHTS

- GIS with big data provides geospatial information to fight COVID-19.
- Big data showed power on epidemic transmission analysis and prevention decision making support.
- Challenges still continue in data aggregation, knowledge discovery, and dynamic expression.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 1 March 2020
Received in revised form 15 March 2020
Accepted 16 March 2020
Available online 20 March 2020

Key words:
COVID-19
Big data
GIS
Spatial transmission
Social management

ABSTRACT

The outbreak of the 2019 novel coronavirus disease (COVID-19) has caused more than 100,000 people infected and thousands of deaths. Currently, the number of infections and deaths is still increasing rapidly. COVID-19 seriously threatens human health, production, life, social functioning and international relations. In the fight against COVID-19, Geographic Information Systems (GIS) and big data technologies have played an important role in many aspects, including the rapid aggregation of multi-source big data, rapid visualization of epidemic information, spatial tracking of confirmed cases, prediction of regional transmission, spatial segmentation of the epidemic risk and prevention level, balancing and management of the supply and demand of material resources, and social-emotional guidance and panic elimination, which provided solid spatial information support for decision-making, measures formulation, and effectiveness assessment of COVID-19 prevention and control. GIS has developed and matured relatively quickly and has a complete technological route for data preparation, platform construction, model construction, and map production. However, for the struggle against the widespread epidemic, the main challenge is finding strategies to adjust traditional technical methods and improve speed and accuracy of information provision for social management. At the data level, in the era of big data, data no longer come mainly from the government but are gathered from more diverse enterprises. As a result, the use of GIS faces difficulties in data acquisition and the integration of heterogeneous data, which requires governments, businesses, and academic institutions to jointly promote the formulation of relevant policies. At the technical level, spatial analysis methods for big data are in the ascendancy. Currently and for a long time in the future, the development of GIS should be strengthened to form a data-driven system for rapid knowledge acquisition, which signifies that

* Corresponding authors: Tel.: +86 010-64888956
E-mail addresses: zhouch@lreis.ac.cn (C. Zhou), sufz@lreis.ac.cn (F. Su).
† These authors contributed equally to this work.

https://doi.org/10.1016/j.geosus.2020.03.005

GIS should be used to reinforce the social operation parameterization of models and methods, especially when providing support for social management.

1. Background

The outbreak of 2019 novel coronavirus disease (COVID-19) is a public health emergency of international concern that had caused more than 100,000 infections and 3,830 deaths in more than 100 countries by March 8, 2020 (NHC, 2020; WHO, 2020), seriously affecting economic and social development. On February 28, UN Secretary-General Guterres called on governments to take action to do everything possible to control the COVID-19 epidemic (New.cn, 2020).

The United Nations Sustainable Development Goals (SDGs) aim to address social, economic, and environmental issues from 2015 to 2030 and move towards sustainable development (United Nations Development Programme, 2015). The United Nations SDGs contain 17 goals and 169 targets. SDG 3 aims to ensure healthy lives and promote well-being for all at all ages including the specific in SDG 3.3, which aims to end epidemics of AIDS, tuberculosis, malaria and neglected tropical diseases and to combat hepatitis, water-borne diseases and other communicable diseases by 2030 (United Nations Development Programme, 2015). The COVID-19 epidemic directly threatens the achievement of the above goals for health, and also affects the realization of goals for economic and social development. In the context of global environmental changes, the transmission characteristics of the COVID-19 epidemic have not yet been sufficiently recognized (CDC, 2020). Additionally, the acceleration of global urbanization, increased concentration of population, more frequent and complex interactions, and shortage of medical protection in developing countries all increase the difficulties of the prevention and control of COVID-19.

1.1. China's response to COVID-19

At the beginning of the epidemic, the medical and research communities responded quickly. They quickly isolated the new coronavirus, conducted gene sequencing to determine the intermediate host, actively shared data with the international community, and sent three successive expert teams to Wuhan. On January 23, 2020, the Chinese government took decisive measures to lock down the city of Wuhan and to close the external routes to all cities in Hubei Province (State council, 2020). Each province has successively launched a first-level public health response, which has effectively curbed the spread of the epidemic. China has undertaken enormous personal and socioeconomic losses and has won valuable time for the Chinese and for global prevention and control of the epidemic. On February 3, only 10 days after construction, Huoshenshan Hospital, which is a 1000-bed hospital in Wuhan, Hubei, was put into use (China Daily, 2020a). On February 8, Leishenshan Hospital, which has 1600 beds, was completed and put into use (China Daily, 2020b). In the interim, medical staff from all over the country rushed to Hubei to fight the epidemic. On February 12, the local government took the measure of receiving and curing all patients that should be treated, following which the epidemic in Hubei Province reached a turning point and began to decline. During this period, we utilized GIS and spatial big data technology, which have a high degree of scientific and technological display (Zhou et al., 2016), to provide important scientific and technical support to allow the government to judge the epidemic situation and formulate prevention and control measures (Health Commission of Hubei Province, 2020).

1.2. Spatiotemporal development of COVID-19 in China

During December 2019, a cluster of unexplained viral pneumonia cases was detected in Wuhan city, Hubei Province, China. The COVID-19 epidemic subsequently spread to the rest of Hubei, and to other parts

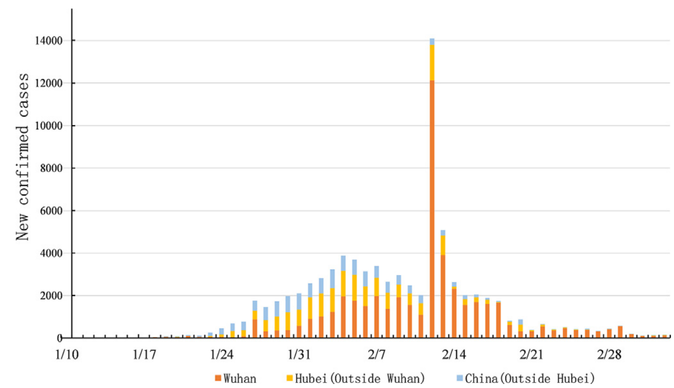


Fig. 1. Daily Changes in new confirmed cases of COVID-19 in China (2020/01/10 - 2020/03/04).
(Data source: National Health Commission of the People's Republic of China. Daily briefing on novel coronavirus cases in China, http://en.nhc.gov.cn/DailyBriefing.html)

of China. The number of newly confirmed cases increased rapidly from January 10 to January 24, 2020, and the reported cases reached a peak and flattened from January 31 to February 7. The number of confirmed cases increased on February 12 because of a change in how cases were diagnosed and reported in Hubei Province that began on the same date. As of March 8, China has cumulatively confirmed 80,735 cases (National Health Commission of the PRC, 2020). The daily confirmed cases from 2020/01/10 to 2020/03/04 are summarized in Fig. 1.

In China, the epidemic outbroke in Wuhan and then spread throughout Hubei Province. Since January 18, because of the large-scale migration associated with the Chinese Lunar New Year, the epidemic has spread rapidly across the country. By January 29, confirmed cases were recorded in all provinces and regions in China. After February 14, the number of newly confirmed cases in the areas outside Hubei Province gradually decreased. By February 21, the number of newly confirmed cases in Hubei Province had increased by one hundred people per day. The newly confirmed cases outside Hubei Province fell to single digits, and the national epidemic situation was effectively controlled. The national distribution maps of newly confirmed cases at the province-level are shown in Fig. 2.

2. Ten challenges in using GIS with spatiotemporal big data

The characteristics of strong infectivity, a long incubation period and uncertain detection of COVID-19, combined with the background of large-scale population flow and other factors, led to the urgent need for scientific and technological support to control and prevent the spread of the epidemic. During the struggle against epidemic, GIS and spatial big data technology have played an important role in identifying the spatial transmission of the epidemic, in spatial prevention and control of the epidemic, in spatial allocation of resources, and in spatial detection of social sentiment, among other things. Here, we discuss ten of these challenges and responses, *viz.*, 1) rapid construction of a big data information system for the epidemic; 2) rapid problem-oriented big data acquisition and integration; 3) convenient multi-scale dynamic mapping for epidemics; 4) comparison between spatial tracking and the spatiotemporal trajectory of big data; 5) spatiotemporal prediction of the transmission speed and scale of the epidemic; 6) spatial segmentation of the epidemic risk and prevention level; 7) spatial dynamic balancing of supply and demand for medical resources; 8) assessment of the supply of materials

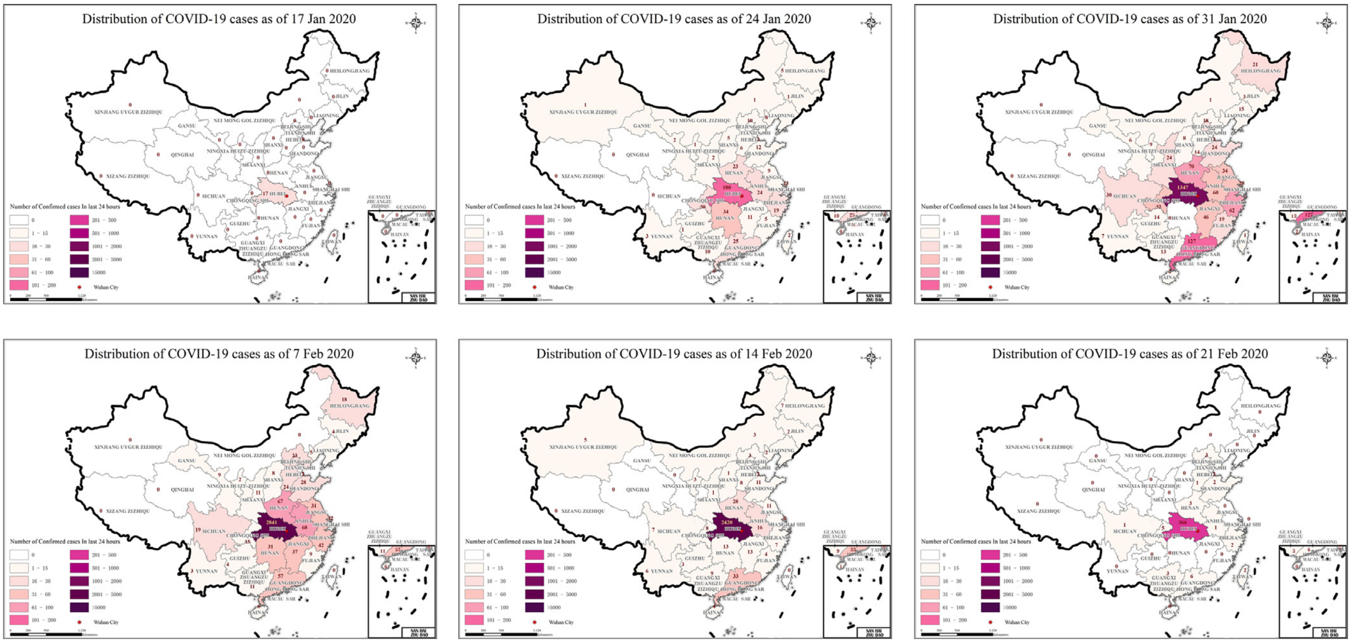


Fig. 2. Changes in new confirmed cases of COVID-19 in China at province-level (2020/01/17 - 2020/02/21).
(Data source: National Health Commission of the People's Republic of China. Daily briefing on novel coronavirus cases in China, http://en.nhc.gov.cn/DailyBriefing.html)

and transportation risk; 9) rapid estimation of the population flow and distribution; and 10) monitoring the spatial spread of social sentiment and detection.

2.1. Rapid construction of a big data information system for epidemics

With the development of GIS technology, an information system for a relevant subject can be constructed rapidly, especially in terms of database management, spatial analysis tools, and mapping. However, the constructed information system is commonly limited by the basic functions of the commercial software. In response to the epidemic, many institutions and research groups have built a number of information systems, such as "epidemic map displays", "fever clinic queries" and "passenger information queries", based on existing commercial software, which has made an important contribution to epidemic prevention and control (CAICT: China Academy of Information and Communications Technology, 2020). Considering that decision-making in regard to epidemic prevention and control requires rapid analysis of the spatiotemporal dynamics and comprehensive consideration of multiple geographical scales, the systems development teams have 1) connected health departments and the internet to build a virtual perception network of multi-source spatiotemporal big data about the epidemic, and developed GIS for the epidemic from a daily time scale, which is a relatively static information system, into real-time dynamic GIS on hourly or even minute-by-minute timescales; 2) built a spatiotemporal cube model of big data about the epidemic, and realized the normalized modeling of multi-source heterogeneous data with different spatial references, different times, different scales and different semantics, as well as creating a unified storage system and management of mixed polymorphic data; 3) set up a computing engine for epidemic description, diagnosis, prediction and decision-making, and developed the traditional online GIS for epidemics with the "visualizing query" function to an integrated stage with "visualization query analysis"; 4) developed a multi-scale integrated spatiotemporal dynamic visualization technology to visualize the epidemic at the "country, province, city, county, community, and individual" scales in order to manage visualization analysis on "one map" of multidimensional epidemic data under a unified spatiotemporal datum; 5) adopted the new generation of native cloud architecture technology to develop a three-tier structure consisting of the infrastructure as the background, a platform for spatiotemporal big data management as the middle ground, and application for the epidemic as the foreground, which solved problems caused by traditional information systems, such as overfull links in development mode, complex processing and long lead time, and satisfied the demand for rapid construction of GIS for an epidemic in an emergency. The system interfaces at different scales are shown in Fig. 3.

2.2. Rapid problem-oriented big data acquisition and integration

The decisions and actions of large-scale epidemic prevention and control depend on data support. The development and application of big data will undoubtedly contribute to quickly identifying the spatiotemporal process of epidemic development, prevention and control measures, and the resulting effectiveness. Strategies for gathering and integrating massive geographic and social-spatial information in the face of prevention and control of an epidemic emergency is the most basic problem for subsequent temporal and spatial mining and analysis. Based on the unified geographical framework, this research quickly absorbed and integrated geographical big data, including the WHO data released internationally, daily domestic health and disease control data, professional population health platform data, Tencent location request data, Baidu migration data, microblog text data, patient spatiotemporal trajectory data, international airline data, census data, education enrollment data, land cover data, remote sensing imagery and other multi-source data through statistical data collection, network data mining, API interface connection, and international and domestic data platform exchange. With the support of unified spatial registration, raster-vector transformation, statistical normalization, format transformation, etc., these data were rapidly merged and applied to the spatiotemporal analysis and visualization systems for the COVID-19 epidemic. In this process, due to the weakness of current research on the classification and security management of spatiotemporal big data, it is clear that GIS needs to be developed in the aspects of automatic correlation aggregation of data,

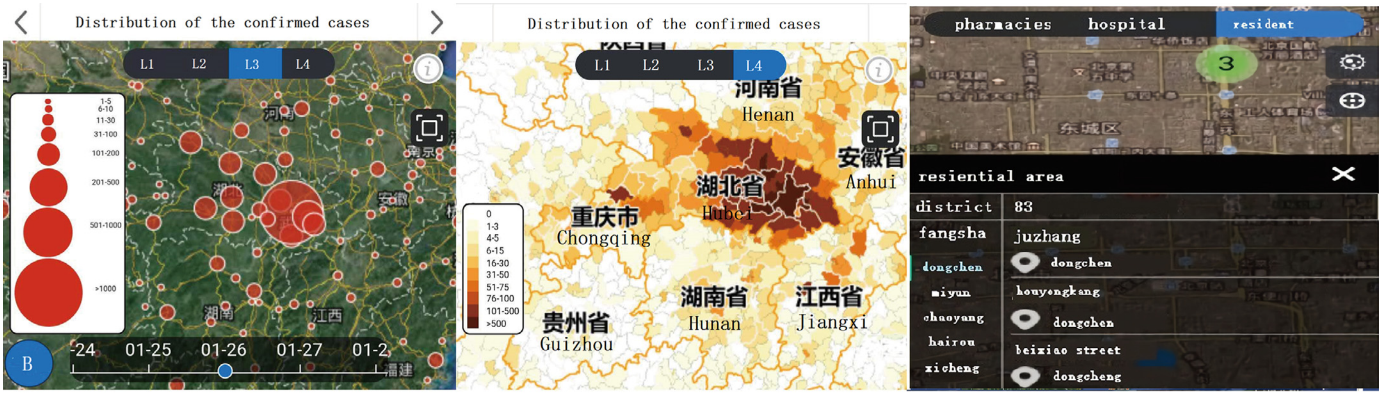


Fig. 3. Dynamic information query system for different scales: (a) City level; (b) County level; (c) Community level.

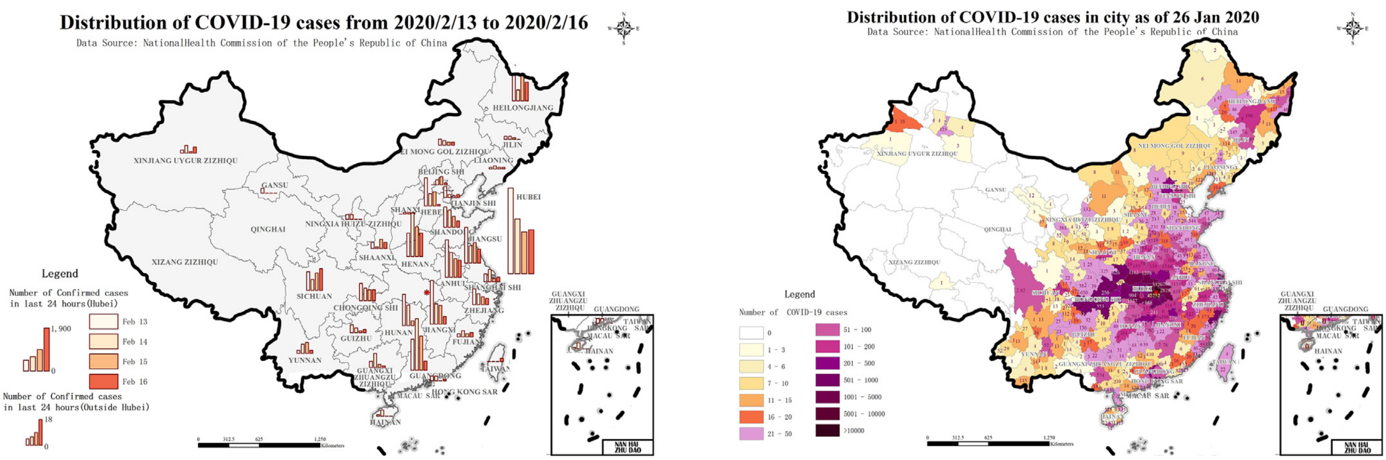


Fig. 4. Rapid mapping based on multi-scale templates.

historical data tracing, adaptive conversion of heterogeneous data, and standardized storage of multi-source data.

2.3. Convenient multi-scale dynamic mapping for epidemics

Since the outbreak of COVID-19, in addition to the news published by the government, epidemic information has also been widely disseminated through internet platforms such as Weibo, WeChat, and other channels. This massive and multi-source information has created considerable challenges for epidemic mapping. ESRI's expert Kenneth Field appealed that coronavirus mapping should be responsible (Field, 2020). He called attention to misunderstandings in the current COVID-19 epidemic maps, including the incorrect use of map projections, choropleth maps, classification schemes, color schemes, point density maps, graduated symbols, heat maps, and three-dimensional maps.

In this study, a data-driven multi-scale mapping template was designed. By rapidly processing real-time public outbreak data in 34 provinces and more than 300 prefecture-level cities, we implemented a multi-scale map matching database template. By making the thematic map templates in advance, the rapid production and release of large-scale epidemic maps were realized. Several color schemes were designed to not only reflect thematic knowledge with the use of color but also to provide more emotively intuitive information, such as a red-purple color scheme for the COVID-19 cases, a blue-black color scheme for the cases with mortality, and a yellow-green color scheme for cured cases. Daily animation maps were applied to express the spatiotemporal characteristics of the spread of the epidemic. The knowledge mapping method combined with the epidemiological and emergency response-related professional knowledge were also utilized further. We designed maps for the multidimensional dynamic expression of the epidemic situation, such as the cumulative distribution map of confirmed cases per 100,000 people and the distribution map of places with zero new recorded cases. We also launched daily reports called the Epidemic Map Story starting on February 1, 2020, via the official WeChat platform. More than ten daily update maps for the public were published and updated, including those for the global COVID-19 epidemic situation, the spatial distribution of the COVID-19 epidemic situation in China, and the development and change of the epidemic situation as well as the spatial distribution and change information of the epidemic situation in critical provinces; there were more than 50,000 page views in total up to March 4, 2020. Two template application examples are shown in Fig. 4.

2.4. Comparison between spatial tracking and spatiotemporal trajectory of big data

The comparison of the spatial track of the patient's activities is a critical technical task for virus tracking and transmission chain reconstruction. The comparison of activity tracks between patients and populations provides an important scientific basis for delimiting the potentially infected population. Methods to quickly and automatically extract the patient's spatiotemporal trajectory from text data, establish the spatiotemporal comparison method, find the potential spatiotemporal exposure link of patients, support epidemiological investigation and rapid analysis, and realize automatic detection of the cross-regional epidemic infection path are major challenges for GIS. In this research, we 1) developed a reconstruction technology for the progress of the spatiotem-

Fig. 5. Exposure analysis of a patient's spatial trajectory.

poral events in a patient's text track data that can automatically convert the track text into quantitative spatiotemporal events; 2) established a spatiotemporal events database with more than 70000 pieces of patient track text covering the whole country; 3) constructed an exposure calculation model and a patient-node-patient association model that integrates time, space and text similarity, which showed the exposure assessment and site risk assessment of each individual (the exposure analysis of a patient's spatial trajectory is shown in Fig. 5). Based on the above work, the key epidemic transmission centers were located, including Baodi Mall in Tianjin; Toulong Mall in Harbin, Heilongjiang; and Yintaishimao Mall in Wenzhou, Zhejiang.

2.5. Spatiotemporal prediction of transmission speed and magnitude

The spatiotemporal spread of infectious diseases in large populations is a very large and complex system that poses great challenges to mathematical modeling (Grassly and Fraser, 2008; Riley, 2007). This research conducted a spatial simulation from the perspective of the geographical environment and social space. A spatiotemporal diffusion model (Multi susceptible-exposed-infectious-removed-died-cumulative model, multi-SEIRDC model) centered on Wuhan was established, which takes into account factors including the effects of spatial barriers from human intervention, the impact of large-scale population migration during the Chinese New Year and the spatial heterogeneity of different epidemic areas. The multi-SEIRDC model was used to track, derive and predict the COVID-19 epidemic situation in different regions of China. Research has shown that the earliest time when COVID-19 began to spread from person to person was in late November or early December. The expected basic reproduction number was 4.08, with a range of 3.37-4.77. The day before the Wuhan lockdown, there were approximately 20,000 people infected with COVID-19 nationwide. As of March 5, the number of potentially infected people in Wuhan exceeds 100,000. Due to the serious lack of detection capacity, there may have been a large number of neglected mild and asymptomatic infections before February 13, 2020. The effective reproduction number outside Hubei fell below the threshold of 1 on February 2, and reached an inflection point and entered a steady decline. The number of new cases per day will fluctuate under 10 cases for approximately one month before the end of the COVID-19 epidemic. If there is no import of overseas infection cases, the end of the COVID-19 epidemic outside Hubei is expected to be at mid-March.

2.6. Spatial segmentation of the epidemic risk and prevention level

Assessing the risk of epidemics and transmission in different regions is of great significance for decision-making and the adjustment of prevention and control efforts. Considering that the epidemic center of China was in Wuhan, the correlation between the number of confirmed cases in each province and the population flow from Wuhan to each province was examined first. The research indicated that until 24:00 on February 2, 2020, the severity of the epidemic in each province was highly correlated with the population who travelled there from Wuhan before the city of Wuhan was locked down, with a correlation coefficient of 0.77. A risk assessment model was constructed with the spatial distribution of the number of confirmed cases and the population migration, and three risk level areas were outlined on the regional scale and on the urban scale for the cities with high risks of epidemics, including Beijing, Shenzhen, Guangzhou, Shanghai, Chongqing, Wenzhou, Zhuhai, Changsha and Harbin. However, after the Spring Festival, the epidemic risk had a high probability of increasing due to the return flow. The three variables, viz., case number, population migration, and transportation network, were incorporated into the prediction. The results (not including those for Hubei) suggested that, after the Spring Festival Holiday, Beijing, Shenzhen, Guangzhou, and Shanghai would have the highest levels of risk, followed by Chongqing, Changsha, Hangzhou, Zhengzhou, Nanjing, Xi'an, and Chengdu. See Fig. 6 for details.

2.7. Spatial dynamic balancing of supply and demand for medical resources

The spatial distribution of medical resources is generally balanced according to factors such as population density, but the spatially uneven outbreak of the epidemic and their rapid development result in a spatiotemporal imbalance of the supply and demand for medical re-

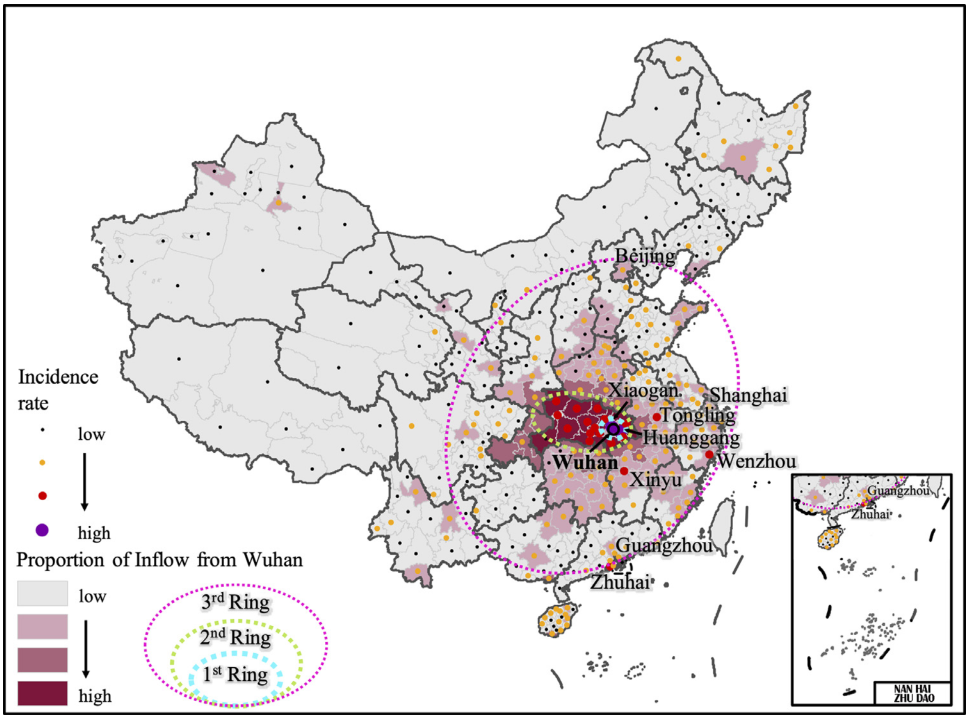


Fig. 6. National spatial segmentation of the COVID-19 epidemic risk.

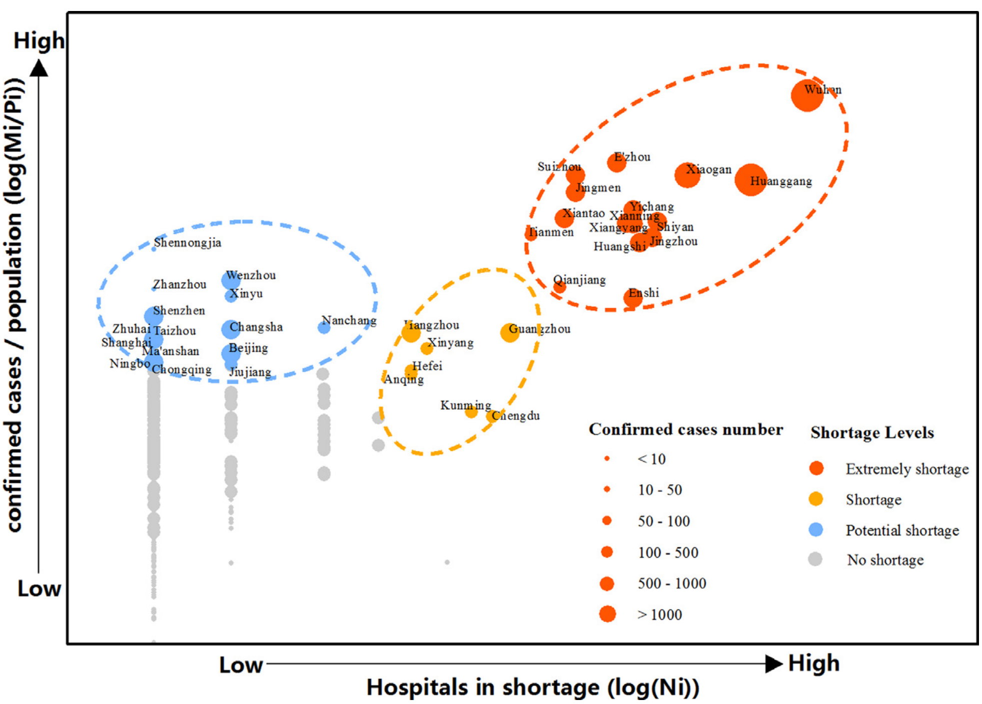


Fig. 7. Relationship between confirmed case number and hospital shortage levels at the city-scale. Mi is cumulative cases number of city i; Pi is estimation number of population of city i; Ni is number of hospitals in shortage of city i; (Statistics until 19:00 2020/02/02).

sources. In this case, the key to epidemic prevention and control is knowing the spatiotemporal dynamics of the supply and demand for medical resources to optimize the allocation of materials. Based on the factors of online hospital help information, local cases and forecasts, and existing resource data, we analyzed the current dynamic situation of medical protective equipment across the country through cross-validation and sampling verification (phone inquiry and web inquiry) and attained the following: 1) the prompt identification of a shortage of medical protective equipments in 462 hospitals, including 336 in Hubei Province and the rest in Sichuan Province, Anhui Province, Guangdong Province, Jiangsu Province and Hunan Province, among others; and 2) based on the number of confirmed cases, the number of hospitals that lacked sufficient medical protective equipments and the urban population, the shortage of medical supplies was divided into four levels on a regional scale and three levels on a city scale (the result for Hubei Province is shown in Fig. 7, and the national result is shown in Fig. 8). The above achievements provided an important scientific basis for the national al-

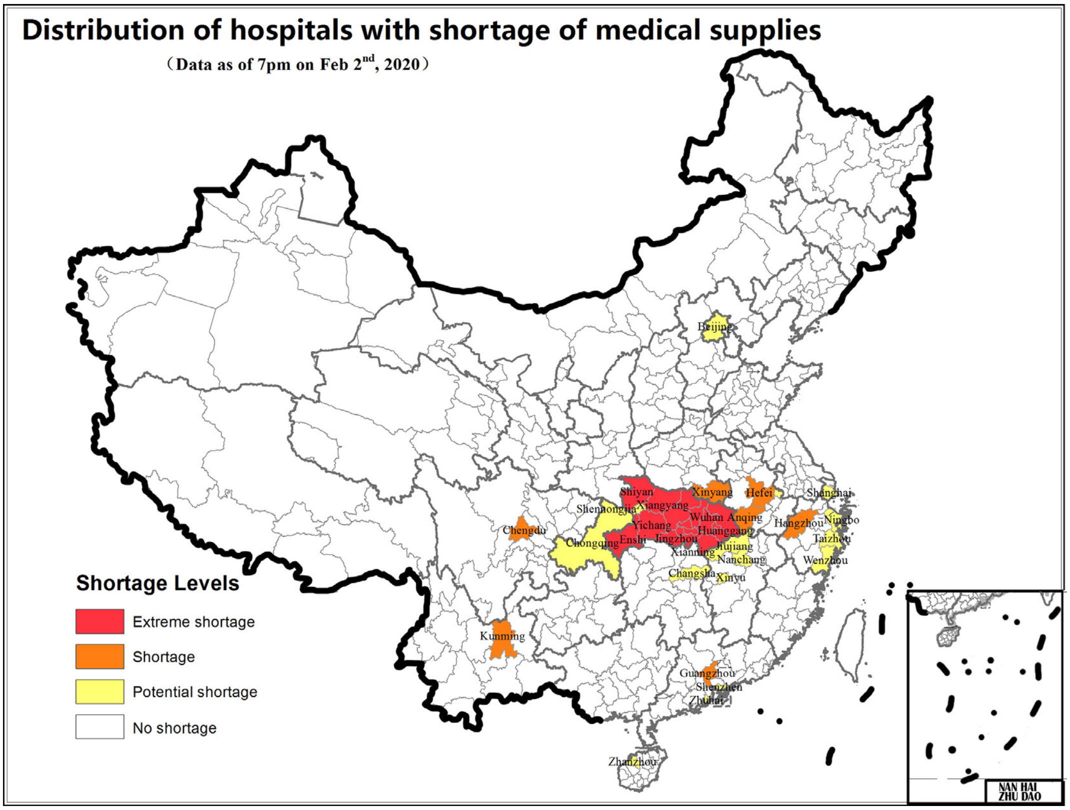


Fig. 8. National distribution of hospitals in shortage of medical protective materials.

location of medical care resources for the prevention and control of the epidemic.

2.8. Assessment of material supply and transportation risk

A stable and efficient national material supply and transportation system provides important support for successful epidemic prevention and control. We have integrated multiple datasets, such as provincial epidemic data, online consumption data and postal service data, to analyze the supply-demand situation and price changes in necessities and food, including vegetables and meat, for every province during the epidemic prevention period, as well as the changing volume and trends of postal and express delivery businesses in each region, to identify the area, type, and the transportation support capacity of the material shortage risks. Through this, we provided scientific data for the social management departments to obtain the material supply-demand dynamic information in real time. Meanwhile, by tracking the transportation of materials, we identified the highly sensitive nodes that may invoke virus transmission during the transportation process and provided advance warning and decision support for the prevention and control of the regional spread of the epidemic (the national distribution of risk index is shown in Fig. 9). Due to the involvement of company business information in the data, there were many difficulties in the acquisition process, which revealed that knowledge of how businesses constrain data sharing will become an important research direction. At present, JD.COM, SF Express, and other large domestic online shopping and logistics companies have begun to establish GIS-based logistics monitoring systems. In the future, with the support of Internet of Things technology, a national classified material transportation monitoring system and national data integration and analysis platform will be gradually established, which will provide more accurate and timely information on the material supply and transportation capacity during emergency policy-making for the whole society.

2.9. Rapid estimation of population flow and distribution

The magnitude and scale of population mobility are essential information for spatial transmission prediction, risk area division, and control measure decision-making for infectious diseases (Wang et al., 2019). Research has shown a strong relationship between the number of train journeys and the number of COVID-19 cases (Zhao et al., 2020). Before the lockdown of Wuhan on January 23, 2020, more than five million people had left Wuhan to other regions in China (News China, 2020). A better understanding of the destinations of those people can assist in decision-making and prevention of coronavirus spread. Therefore, in this research, by using multi-source spatiotemporal big data, including Tencent location request data, Baidu migration data, and land cover data, we have developed a dynamic estimation model of the multilevel spatial distribution of the interregional migrant population and further characterized the spatial distribution of the population migrating from Wuhan to other regions of Hubei Province. The results show that 1) during the Spring Festival, the average ratio between the increases in rural populations to the total population was 124.7% in the prefecture-level cities in Hubei Province, and at least 51.3% of the population moving from Wuhan to prefecture-level cities flowed into rural areas, and 2) the spatial distribution of migrants among cities and counties in Hubei Province exhibited a three-ringed structure. The first ring was the core area of disease, which included Wuhan and its surrounding areas, which mainly experienced population outflows. The second ring was the primary focus area, including Huanggang, Huangshi, Xiantao, Tianmen, Qianjiang, Suizhou, Xiangyang, and parts of Xiaogan, Jingzhou, Jingmen, and Xianning, where the total population and the population in rural areas increased significantly during the Spring Festival. The third ring was the secondary focus area, which included Yichang, Enshi, Shennongjia, and parts of Jingzhou and Jingmen, which were located in the western part of Hubei Province and were mainly characterized by a small population inflow (shown in Fig. 10). The above results were shared with government officials, and the development of the epidemic

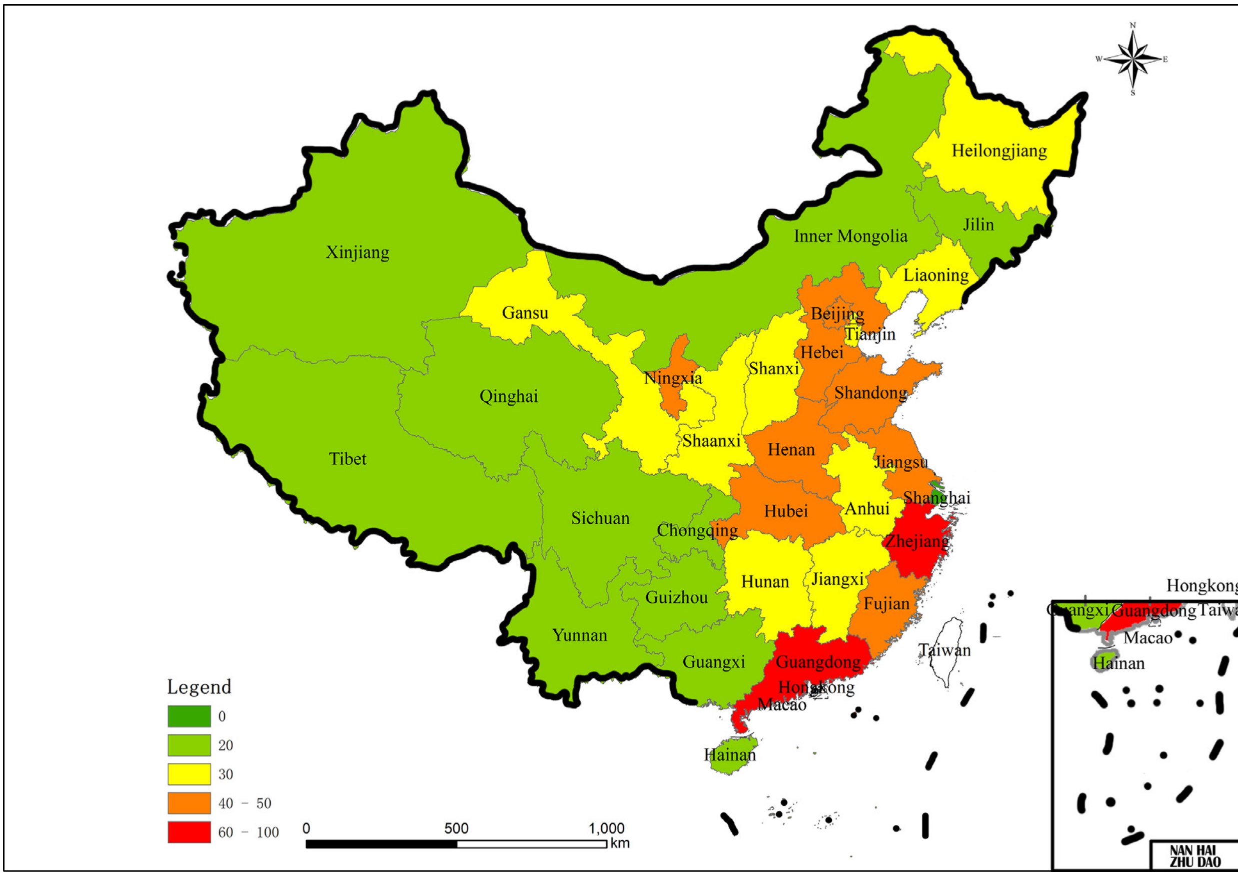


Fig. 9. Risk index of new coronavirus transmission associated with the logistics process in China on January 31, 2020.

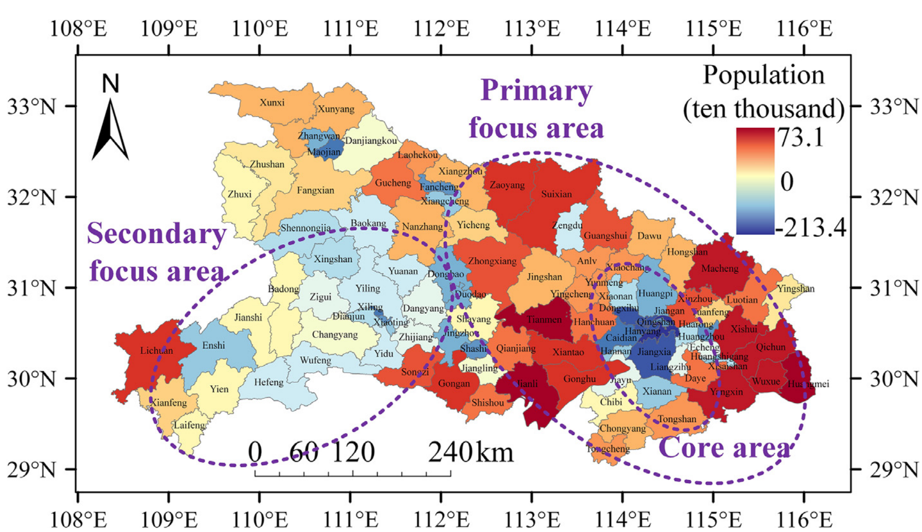


Fig. 10. The spatial distribution of county-level total population changes during the 2018 Spring Festival (Lunar New Year's Day- the fourth day after Lunar New Year's Day) in Hubei Province.

confirmed the objectivity of this research, which strongly supported government decision-making.

Population mobility in the Spring Festival of 2020 was significantly lower than that during the Spring Festival of 2019 (shown in Fig. 11) due to the quarantine measure, which prevented the wider spread of the epidemic (Hu, 2019; Wei et al., 2018). However, Fig. 11 shows that population mobility has increased since February 17 (the 24th of the first lunar month of 2020). The return of people to work will possibly increase the spread of COVID-19. Baidu migration data were utilized here to predict the risk (Xu et al., 2017). We calculated the speed of the population mobility recovery in each city (shown in Fig. 12a.) based on the population migration from February 17-23, 2020. The result showed that the speed of the return to normal levels of population mobility was rapid in economically developed cities in South China. The community division of the population mobility network showed that the population flows were mostly from the cities in Henan, Anhui, and Jiangsu to the Yangtze River Delta and from cities in Hunan, Guizhou, and Jiangxi to the Pearl River Delta. Both the departure cities and the destination cities are places with large outbreak of COVID-19. Based on the above analysis, taking strict measures on these moving populations was suggested in the case of the second surge of COVID-19 cases.

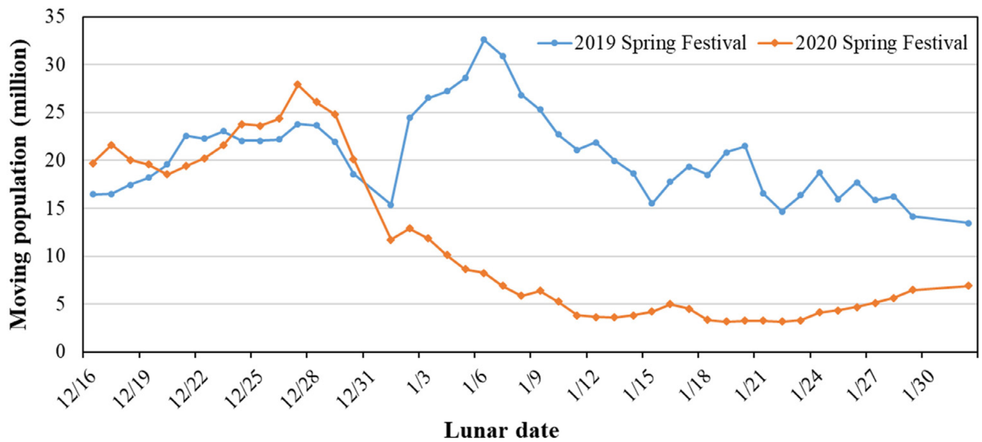


Fig. 11. National population flow trend in the same period of Spring Festival in 2019 and 2020.

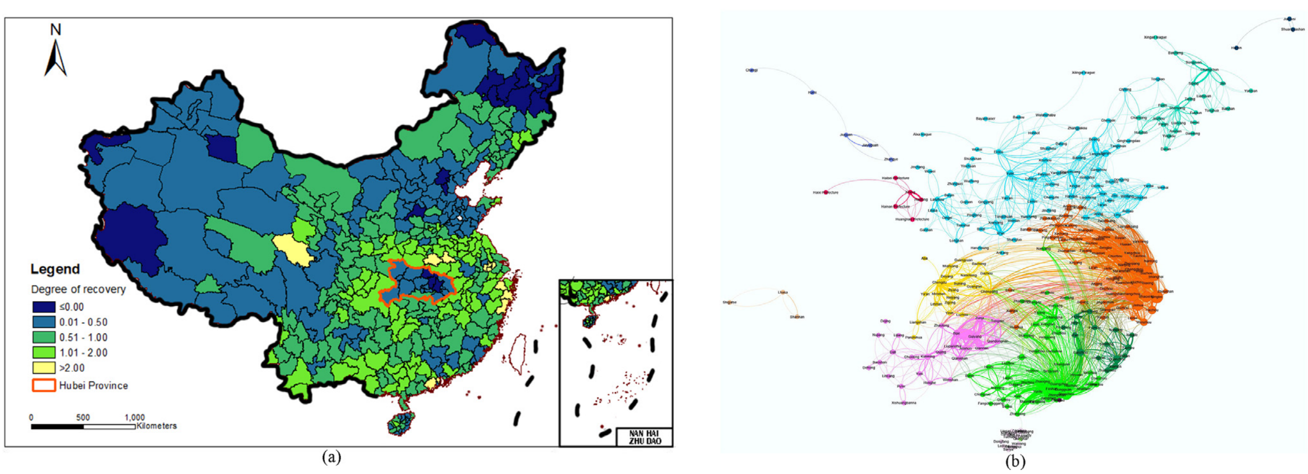


Fig. 12. (a) Recovery of urban population flow; (b) Rework population flow network and community Division.

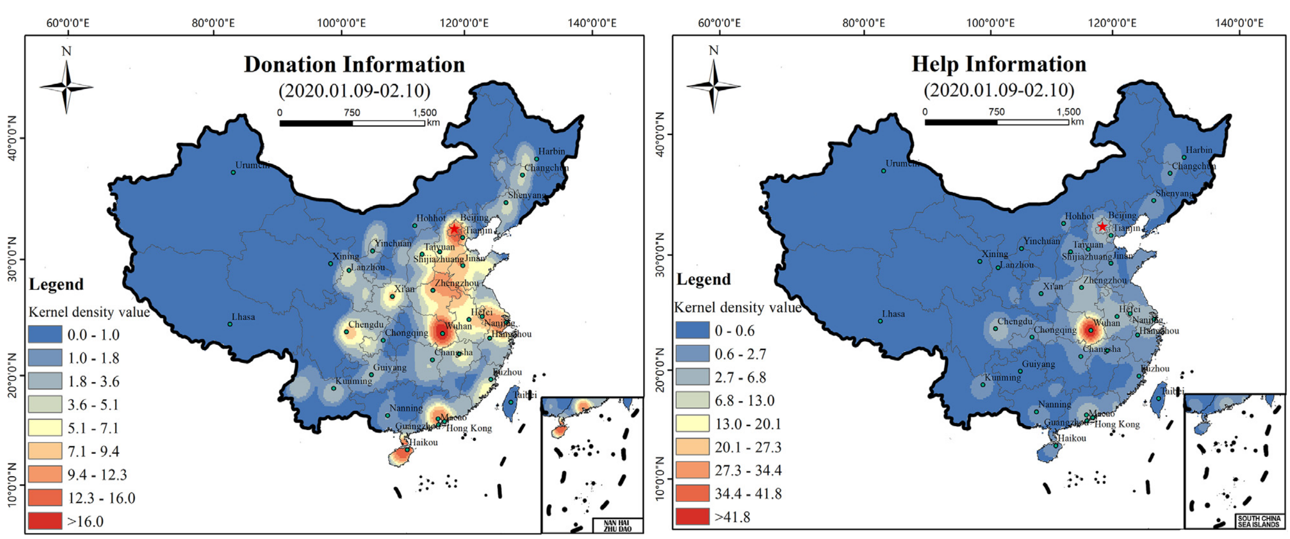


Fig. 13. Spatial distribution of help and donation information of COVID-19 during the epidemic period (2020/01/09 - 2020/02/10).

2.10. Spatial spread and detection of social sentiment

When a major epidemic occurs, the negative impact of uncertainty and panic on social operations may exceed that of the viral diseases. Therefore, this research applied massive social media data to track and evaluate the spatial spread of public sentiment (Miller and Goodchild, 2015). Considering that public behavior has the characteristics of irrationality, strong infectivity and conformity, we needed to build a knowledge base of epidemic sentiment association by mining the dynamic evolution process of public opinion in time and space and using the semantics from social media (Goodchild and Glennon, 2010; Shi et al., 2017). The Sina Weibo, which is similar to Twitter, is the social media platform on which Chinese people typically share their opinions. With the goal of modelling and visualizing the semantic and spatiotemporal evolution of public opinion on COVID-19, a topic extraction and classification framework for extracting topics of a COVID-19-related microblogs was designed and implemented. Based on the complex network, a user semantic behavior evolution model was introduced to measure and analyze the change in public opinion by tracking topics and sentiments that appeared in microblog users' timelines in response to COVID-19. The results indicated that from January 9 to February 10, 2020, more than 60% of posts related to science popularization of disease prevention, government announcements, and responses suggested that public sentiment was positive and stable. The posts with the topics of "help-seeking" were concentrated in the key epidemic area of Wuhan, and the posts related to "donation information" were widely distributed throughout the country (shown in Fig. 13). The results reflected the characteristics of China's emergency disaster relief, which was characterized by the adage "Trouble on one side, help from all sides".

3. Conclusions

COVID-19 is characterized by a long incubation period, strong infectivity and difficulty of detection, which has led to the sudden outbreak and the rapid development of an epidemic. This situation requires GIS and big data technology to allow rapid responses and analyses, a quick supply of information about the epidemic dynamics and an understanding of the epidemic development rules to provide timely support for the prevention and control decisions and actions.

In this study, we analyzed the spatial representation of the disease, material, population and social psychology at three scales: individual, group and regional. At the individual scale, the comparison between spatial epidemic tracking and the spatiotemporal trajectories of patients was carried out. At the group scale, the estimation of population flow and the spatial distribution was carried out. At the regional scale, the segmentation of spatial risk, the analysis of balance between the supply and demand of medical resources, and the spatial differentiation analysis of material transportation capacity and social sentiment were carried out.

From the perspective of GIS technology, this study revolutionized the data acquisition methods through big data technology, which achieved rapid data acquisition and integration from traditional data to big data of various organizations. The analysis platform was quickly constructed through an innovative construction technology system, which provided the technical platform for timely epidemic analysis. The production of epidemic maps was quickly completed through multi-scale dynamic template technology, which allowed timely dissemination of epidemic dynamic information. From the perspective of spatial simulation and analysis, this study simulated the spatial transmission process of the epidemic scientifically by increasing detailed regional variables and changing the population flow and R_0 (basic reproduction number) according to the prevention and control measures. The calculation of the overlap between the spatial tracking of the virus and that of patient trajectories was realized based on the exposure index of fused spatial and text information via text spatialization technology. The estimation of population flow and spatial distribution was realized via the combination of big data and traditional geographic data, which identified the problems in the key risk areas and the spatial mismatching of medical resources in a timely manner, to bring about a rapid supply of information about the delimitation of prevention areas and resource deployment in epidemic control. The monitoring of social sentiment from social media was realized through the construction of a knowledge database on public opinion, which provided important foundational information on public opinion to guide the government.

In assessing the contribution of GIS and spatial big data technology to the containment of the COVID-19 epidemic, it is clear that many challenges remain to be studied. For example, the status of big data source restrictions in commercial enterprises may restrict the data supply needed for social management, result in the lack of a mature scheme for big data aggregation, and cause difficulty of rapid online application of deep integration, which are ongoing issues. Regarding data-driven knowledge acquisition, the uncertainty of social operations, especially with the high spatial heterogeneity of responses to epidemic development throughout the country, may lead to spatial deviations from the model simulation. Strategies for a technical system of knowledge acquisition based on big spatial data focused on social operations are existing challenges. From the aspect of result presentation, the status presentation can be fully reflected, while the multi-scale dynamic presentation driven by big data is urgently needed and promising.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgment

This work was funded by the National Natural Science Foundation of China (41421001, 42041001 and 41525004). We thank the reviewers who provided valuable comments on this paper. We are very grateful to those who participated in the data preparation, processing and analysis for this study. We ackonwledge Professor Michael E Meadows and colleagues in International Geographic Union and in its Commission on Geographical information Science. We also acknowledge all the support from the State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences.

References

CAICT: China Academy of Information and Communications Technology, 2020. Research Report on Data and Intelligent Application in Epidemic Prevention and Control (1.0). http://www.caict.ac.cn/kxyj/qwfb/ztbg/202003/P020200305495005485729.pdf (accessed 01 March 2020). (in Chinese)
CDC: Centers for Disease Control and Prevention, 2020. Healthcare Professionals: Frequently Asked Questions and Answers. https://www.cdc.gov/coronavirus/2019-ncov/hcp/faq.html/ (accessed 21 February 2020).
China Daily, 2020a. Superfast hospital in Wuhan ready. China Daily. https://epaper.chinadaily.com.cn/a/202002/03/WS5e37713ca310a2fabb7a1dc7.html/ (accessed 02 February 2020).
China Daily, 2020b. Wuhan: Leishenshan Hospital Undergoing Quality Check Before Phased Delivery. China Daily. https://www.chinadaily.com.cn/a/202002/08/WS5e3e7810a310128217275fb3.html/ (accessed 02 February 2020).
Field, K, 2020. Mapping coronavirus, responsibly. https://www.esri.com/arcgis-blog/products/product/mapping/mapping-coronavirus-responsibly/ (accessed 25 February 2020).
Goodchild, M., Glennon, J., 2010. Crowdsourcing geographic information for disaster response: a research frontier. Int. J. Digital Earth 3 (3), 231–241.
Grassly, N., Fraser, C., 2008. Mathematical models of infectious disease transmission. Nat. Rev. Microbiol. 6 (6), 477–487.
Health Commission of Hubei Province, 2020. http://wjw.hubei.gov.cn/bmdt/ztzl/fkxxgzbdgrfyyq/xxfb/202002/t20200213_2025580.shtml/ (accessed 26 January 2020). (In Chinese)
Hu, M., 2019. Visualizing the largest annual human migration during the Spring Festival travel season in China. Environ. Plann. A: Econ. Space 51 (8), 1618–1621.
Miller, H., Goodchild, M., 2015. Data-driven geography. Geojournal 80 (4), 449–461.
National Health Commission of the PRC, 2020. COVID-19 epidemic situation up to 24:00 on March 8th. http://www.nhc.gov.cn/xcs/yqtb/202003/f2c83db9f73d4be5be0dc96af731813c.shtml/ (accessed 9 March 2020).
New.cn, 2020. COVID-19 is urged by the UN Secretary General to do everything possible to contain the outbreak. http://www.xinhuanet.com/2020-02/29/c_1125642849.htm/ (accessed 29 February 2020).
News China, 2020. Mayor of Wuhan: more than 5 million people left Wuhan. http://news.china.com.cn/2020-01/26/content_75650784.htm/ (accessed 26 January 2020) (In Chinese).
Riley, S., 2007. Large-Scale Spatial-Transmission Models of Infectious Disease. Science 316 (5829), 1298–1301.
Shi, X., Ma, L., Li, B., 2017. Study on disaster information management system compatible with VGI and crowdsourcing. Sci. Surv. Mapp. 42 (3), 191–195. (in Chinese)
State council, 2020. http://www.gov.cn/xinwen/2020-01/23/content_5471751.htm/ (accessed 23 January 2020). (in Chinese)
United Nations Development Programme, 2015. What are the Sustainable Development Goals. https://www.undp.org/content/undp/en/home/sustainable-development-goals.html/ (accessed 26 January 2020).
Wang, Y., Dong, L., Liu, Y., Huang, Z., Liu, Y., 2019. Migration patterns in China extracted from mobile positioning data. Habitat Int. 86, 71–80.
Wei, Y., Song, W., Xiu, C., Zhao, Z., 2018. The rich-club phenomenon of China's population flow network during the country's spring festival. Appl. Geogr. 96, 77–85.

WHO: World Health Organization, 2020. Coronavirus disease (COVID-2019) situation Reports. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/ (accessed 21 January 2020).

Xu, J., Li, A., Li, D., Liu, Y., Du, Y., Pei, T., et al., 2017. Difference of urban development in China from the perspective of passenger transport around Spring Festival. Appl. Geogr. 87, 85–96.

Zhao, S., Zhuang, Z., Ran, J., Lin, J., Yang, G., Yang, L., He, D., 2020. The association between domestic train transportation and novel coronavirus (2019-nCoV) outbreak in China from 2019 to 2020: A data-driven correlational report. Travel Med. Infect. Dis. 33, 101568.

Zhou, C., Su, F., Harvey, F., Xu, J., 2016. Spatial Data Handling in Big Data Era. Springer, Beijing.